

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО**

Факультет інформатики та обчислювальної техніки  
(назва факультету, інституту)

Кафедра автоматизованих систем обробки інформації і управління  
(назва кафедри)

"На правах рукопису"

УДК 519.68; 681.513.7;

612.8.001.57; 007.51/.52

«До захисту допущено»

Завідувач кафедри

О.А.Павлов

(підпис)

(ініціали, прізвище)

“ ” 20 18 р.

**МАГІСТЕРСЬКА ДИСЕРТАЦІЯ**

**на здобуття ступеня магістра**

за спеціальністю 122 Комп'ютерні науки та інформаційні технології  
(код та назва спеціальності)

спеціалізацією Інформаційні управляючі системи та технології  
(код та назва спеціалізації)

на тему: Прогнозування результатів спортивних змагань за допомогою  
марківських моделей та нейронних мереж

Виконала: студентка VI курсу групи ІС-61м  
(шифр групи)

Дідківська Валерія Андріївна

(прізвище, ім'я, по батькові)

(підпис)

**Науковий керівник** к.ф.-м.н., доц. Гавриленко О.В.

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

**Консультант** к.т.н., доц. Жданова О.Г.

(науковий ступінь, вчене звання, прізвище, ініціали)

(підпис)

**Рецензент**

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Засвідчую, що у цій магістерській дисертації  
немає запозичень з праць інших авторів без  
відповідних посилань.

Студент

(підпис)

Київ – 2018





## РЕФЕРАТ

Магістерська дисертація: 119 с., 20 табл., 2 додатки, 50 джерел.

**Актуальність.** Прогнозування результатів спортивних змагань – область досліджень, яка розвивається не лише за рахунок постійно зростаючого інтересу букмекерів, а й для ґрунтовнішої підготовки спортсменів. За допомогою прогнозування результатів змагань можна визначити “слабкі” місця спортсмена у кількісних значеннях, визначити мету, якої варто досягти до наступних змагань та запропонувати шляхи або засоби, за допомогою яких це досягнення може стати можливим.

У зв’язку з цим науковою задачею є розробка нового методу прогнозування результатів спортивних змагань, який дозволив би ефективніше розв’язувати задачу прогнозування, комбінуючи наявні та модифіковані методи.

**Зв’язок роботи з науковими програмами, планами, темами.** Робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «Інтелектуальний аналіз даних» (№УДК 519.68; 681.513.7; 612.8.001.57; 007.51/.52).

**Мета і завдання дослідження** – підвищення ефективності прогнозування результатів спортивних матчів.

Для досягнення мети необхідно виконати наступні **задачі**:

- розробити формальну постановку задачі прогнозування спортивних змагань;
- розробити новий системний підхід до прогнозування результатів спортивних змагань;
- розробити методологію прогнозування результатів спортивних змагань, яка поєднує в собі метод прогнозування ланцюгами Маркова та нейронними мережами;
- модифікувати алгоритм прогнозування результатів спортивних змагань за допомогою ланцюгів Маркова для прогнозування спортивних змагань;

- впровадити методологію прогнозування спортивних змагань у вигляді програмного продукту, який може використовуватися для прогнозування.

**Об'єкт дослідження** – процес прогнозування результатів спортивних змагань.

**Предмет дослідження** – методи та моделі прогнозування результатів спортивних змагань.

**Методи дослідження**, застосовані у даній роботі, базуються на методах машинного навчання та теорії ймовірності.

**Наукова новизна одержаних результатів** полягає у розробці принципово нової методології прогнозування результатів спортивних змагань, яка включає в себе алгоритми на основі ланцюгів Маркова та нейронних мереж. Поєднання цих методів дозволило проводити ретельний аналіз даних та використовувати модифіковані методи при прогнозуванні.

**Апробація результатів.** Тези доповіді на 8-й міжнародній науково-технічній конференції “Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління”, м. Харків. Тези доповіді на 4-й міжнародній науково-практичній конференції “Актуальні питання сучасної науки, м. Київ”. Тези доповіді на науково-практичній конференції “Інформатика та обчислювальна техніка ІОТ-2018”, м. Київ. Випуск Стартап Школи “Sikorsky Challenge”.

**Публікації.** За матеріалами дисертації було опубліковано наукову статтю в науково-технічному збірник “Автомобільні дороги і дорожнє будівництво”, Серія “Технічні науки” (ISSN 0365-8171).

ПРОГНОЗУВАННЯ СПОРТИВНИХ ЗМАГАНЬ, ЛАНЦЮГИ МАРКОВА, ТЕОРІЯ ЙМОВІРНОСТІ, ЙМОВІРНІСТЬ ВИГРАШУ, НЕЙРОННІ МЕРЕЖІ, СТАТИСТИКА

## ABSTRACT

Master's thesis: 111 pages, 20 tables, 2 appendixes, 50 references.

**Topicality.** Forecasting the results of sports events is an area of research that develops not only at the expense of the ever-increasing interest of bookmakers, but also for more thorough training of athletes. By predicting the results of the competition, you can determine the "weak" places of the athlete in quantitative terms, determine the goal to be achieved before the next competition and suggest ways or means by which this achievement may become possible.

In this regard, the scientific task is the development of a new method for forecasting the results of sports competitions, which would more effectively solve the problem of prediction by combining existing and modified methods.

### **The work communication with academic programs, plans, themes.**

The work was carried out at the Department of Automated Systems for Information Processing and Management of the National Technical University of Ukraine "Kyiv Polytechnic Institute. Igor Sikorsky "within the framework of the theme" Intellectual Data Analysis "(№UDK 519.68; 681.513.7; 612.8.001.57; 007.51 / .52).

**Aim of the research** is to increase the effectiveness of forecasting the results of sports matches.

### **Tasks of the research.**

To achieve the goal you need to accomplish the following tasks:

- to develop a formal statement of the task of forecasting sporting events;
- develop a new system approach to forecasting the results of sporting events;
- develop a methodology for forecasting the results of sporting events, which combines the method of forecasting the Markov chains and neural networks;
- to modify the algorithm of forecasting the results of sports competitions using the Markov chains for forecasting sports competitions;
- implement a methodology for forecasting sports competitions in the form of a program product that can be used for forecasting.

**Object of the research** is the process of forecasting the results of sporting events.

**Subject of research** - methods and models of forecasting of results of sporting competitions.

**Research methods** used in this paper are based on the methods of machine learning and probability theory.

**Scientific novelty of the results** obtained is the development of a fundamentally new methodology for forecasting the results of sports competitions, which includes algorithms based on Markov chains and neural networks. The combination of these methods allowed for careful data analysis and the use of modified methods for prediction.

**Aprovation of results of thesis.** Abstracts of the report at the 8th international scientific and technical conference "Modern trends in the development of information and communication technologies and means of management", Kharkiv. Abstracts at the 4th International Scientific and Practical Conference "Actual Issues of Modern Science, Kyiv". Abstracts of the report at the scientific and practical conference "Informatics and Computing IOT-2018", Kyiv. Issue of the Startup of the Sikorsky Challenge School.

**Publications.** According to the dissertation materials, a scientific article was published in the scientific and technical collection "Automobile roads and road construction", Series "Technical sciences" (ISSN 0365-8171).

FORECASTING SPORT COMPETITIONS, MARKOV CHAIN, THEORY OF THE POSSIBILITY, THE POSSIBILITY OF THE WINNING, NEURAL NETWORKS, STATISTICS

## ЗМІСТ

РЕФЕРАТ.....	3
ABSTRACT.....	5
ВСТУП.....	10
<b>1 ОГЛЯД ПІДХОДІВ ТА ОСНОВНІ ПОЛОЖЕННЯ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ СПОРТИВНИХ ЗМАГАНЬ .....</b>	<b>13</b>
1.1 Задача прогнозування результатів спортивних змагань .....	13
1.2 Основні положення та елементи турнірів великого тенісу.....	15
1.3 Додаткові правила, техніка та тактика гри, традиції екіпіровки.....	19
1.4 Тенісні турніри .....	22
1.5 Дані для аналізу.....	24
1.6 Збір вхідних даних .....	27
1.7 Кластеризація при передбаченні результатів змагань.....	28
1.8 Відомі результати аналізу факторів результативності гравців.....	30
1.9 Відомі результати дослідження прогнозування результатів спортивних змагань.....	32
1.10 Висновок до розділу .....	33
<b>2 ІСНУЮЧІ ПІДХОДИ ДО ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ СПОРТИВНИХ ЗМАГАНЬ.....</b>	<b>34</b>
2.1 Марківські ланцюги для прогнозування результатів тенісних турнірів.....	34
2.2 Моделювання тенісного гейму .....	36
2.3 Моделювання тенісного тай-брейку .....	38
2.4 Моделювання тенісного сету .....	39
2.5 Моделювання тенісного матчу .....	41
2.6 Оцінка ймовірності виграшу при подачі .....	41
2.7 Машинне навчання .....	42
2.8 Нечітка логіка .....	47
2.9 Задача класифікації .....	48
2.10 Висновок до розділу .....	50
<b>3 РОЗРОБКА МЕТОДУ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ СПОРТИВНИХ ЗМАГАНЬ.....</b>	<b>52</b>
3.1 Ідея запропонованого алгоритму.....	52
3.2 Формальна постановка задачі .....	55
3.3 Модифікований метод ланцюгів Маркова у прогнозуванні тенісного матчу .....	56
3.3.1 Застосування ланцюгів Маркова в ході алгоритму .....	57
3.3.2 Математична постановка .....	57
3.3.3 Допустимі стани та множина пошуку рішень .....	58
3.3.4 Алгоритм проходження по ланцюгу гейму .....	59
3.4 Прогнозування тенісного матчу за допомогою нейронних мереж .....	60
3.4.1 Вплив покриття корту на результат прогнозування.....	60



3.4.2	Застосування нечіткої логіки в процесі прогнозування.....	61
3.4.3	Застосування нейронних мереж в процесі прогнозування.....	64
3.4.4	Застосування логістичної регресії та гіперболічної функції дотичної активзації в процесі прогнозування.....	65
3.4.5	Ідея роботи методу із застосуванням нейронної мережі.....	67
3.5	Алгоритм фінального прогнозування для гравця.....	68
3.5.1	Алгоритм вибору ваг факторів.....	69
3.5.2	Фінальне прогнозування.....	70
3.6	Висновок до розділу.....	71
4	ОПИС ПРОГРАМНОГО ПРОДУКТУ.....	72
4.1	Структура компонент програмного забезпечення.....	72
4.2	Структура діаграми класів.....	73
4.3	Структура діаграми бази даних.....	75
4.4	Структура use-case діаграми.....	76
4.5	Керівництво користувача.....	77
4.6	Висновок до розділу.....	79
5	РОЗРОБКА СТАРТАП-ПРОЕКТУ.....	80
5.1	Резюме стартапу.....	80
5.1.1	Назва проекту.....	80
5.1.2	Ідея проекту.....	80
5.1.3	Партнери проекту.....	80
5.1.4	Технічні партнери проекту.....	81
5.2	Технологічний аудит ідеї проекту.....	81
5.2.1	Розвиток напрямку.....	81
5.2.2	Порівняння з аналогами.....	82
5.2.3	Просторова модель та подавлення перебивок багатоканальності.....	83
5.2.4	Склеювання зображення по калібрувальним тестам.....	84
5.2.5	Автокалібрування.....	87
5.3	Застосування результатів досліджень, проведених в дисертації, для побудови схеми розміщення сенсорів.....	88
5.4	Аналіз ринкових можливостей запуску стартап-проекту.....	89
5.4.1	Поточний стан ринку продукту.....	89
5.4.2	Рішення.....	89
5.5	Розроблення ринкової стратегії.....	90
5.5.1	Історія впровадження.....	90
5.5.2	Ринкова стратегія.....	90
5.5.3	Огляд конкурентів.....	91
5.5.4	Технологічні переваги.....	92
5.6	Прототипи продукту.....	92
5.7	Побудова бізнес-моделі.....	94
5.7.1	Бізнес-модель.....	94
5.7.2	Прогноз розвитку бізнесу.....	95
5.7.3	Досягнення.....	95
5.8	Висновок до розділу.....	95

6	ЕКСПЕРИМЕНТИ .....	97
6.1	Застосування модифікованого методу Маркова .....	97
6.2	Дослідження ефективності алгоритму прогнозування результатів спортивних змагань .....	100
	ПЕРЕЛІК ПОСИЛАНЬ .....	104
	ДОДАТОК А .....	112
	Сертифікат навчання у Стартап Школі “Sikorsky Challenge” .....	112
	ДОДАТОК Б Графічний матеріал .....	114
	ПЛАКАТ 1 Фрагмент діаграми класів .....	115
	ПЛАКАТ 2 Марківський ланцюг для ситуації сету і тай-брейку .....	116
	ПЛАКАТ 3 Діаграма бази даних .....	117
	ПЛАКАТ 4 Діаграма Use-case .....	118
	ПЛАКАТ 5 Схема роботи алгоритму в загальному вигляді .....	119
	ПЛАКАТ 6 Система виводу задачі .....	120
	ПЛАКАТ 7 Екранні форми початку роботи програми .....	121
	ПЛАКАТ 8 Екранні форми процесу роботи програми .....	122

## ВСТУП

**Актуальність теми.** Прогнозування результатів спортивних змагань – область досліджень, яка розвивається не лише за рахунок постійно зростаючого інтересу букмекерів, а й для ґрунтовнішої підготовки спортсменів. За допомогою прогнозування результатів змагань можна визначити “слабкі” місця спортсмена у кількісних значеннях, а також визначити мету, якої варто досягти до наступних змагань та запропонувати шляхи або засоби, за допомогою яких це досягнення може стати можливим.

Однак не менший інтерес до моделювання спортивних змагань виникає у людей, безпосередньо задіяних у цих масових видовищних подіях: тренерів, спортсменів. Моделювання дає змогу оцінити результати під, можливо, неочевидним кутом та допомагає визначитися у подальших побудовах стратегій гри.

Якщо розглядати задачу прогнозування як спосіб отримання метрик для спортсмена, які він може використовувати в подальшому, то в залежності від цілі прогнозування можна виділити короткострокове, середньострокове, довгострокове і наддовге прогнозування.

Короткострокове прогнозування, як правило, створюється для планування окремого заняття спортсмена і направлено на передбачення можливості виконати поставлену ціль. Середньострокове прогнозування допомагає визначити найбільш вірогідні тенденції та темпи росту тренуваності з урахуванням засобів, які можуть бути для цього використані: методи тренувань, системи побудови тренувань тощо. Довгострокове і наддовгострокове прогнозування проводяться задля оптимізації процесу спортивного відбору та отримання прогнозів на наступні 3-4 роки. У розрізі даної роботи результати прогнозування можуть бути використані для побудови методик та систем тренування гравця.

Методологія, створена в рамках даної роботи, може бути використана для аналізування видів спорту, які мають ієрархічну структуру. Використання може бути можливим завдяки впровадженню в систему правил та значень, які відповідають

матчу виду спорту, який має бути проаналізований. Як приклад, у даній роботі було взято за основу великий теніс: його правила, виключні ситуації, побудову гри.

Було розглянуто прогнозування результатів спортивних змагань на прикладі тенісу. Математичне моделювання тенісу набирає популярність на наших очах. Кожен рік з'являються нові аналітичні моделі та сервіси, які змагаються один із одним у точності прогнозування результатів тенісних матчів. Це викликано бажанням заробити на стрімко зростаючому ринку онлайн-спортивних ставок - часто сума ставок на окремий матч у професійному тенісі сягає мільйонів доларів.

Великий теніс – це вражаюче видовище, яке має мільйони шанувальників чи не з усіх частин світу. Асоціація тенісистів-професіоналів (АТР) щорічно проводить більше 60 професійних турнірів у 30 країнах. Потенційний прибуток і науковий інтерес зумовили сплеск досліджень в області алгоритмів точного прогнозування тенісних матчів [17].

Система очок в тенісі має ієрархічну структуру: матч складається з сетів, які складаються з геймів, які складаються з окремих очок. У більшості сучасних підходів до прогнозування тенісу ця структура використовується для отримання ієрархічних виразів ймовірності перемоги гравці у матчі на основі марківських ланцюгів. Якщо вважати, що очки в тенісі мають незалежний однаковий розподіл, то для отримання виразу необхідно знати тільки ймовірність виграшу кожним гравцем очки при подачі. На підставі цієї базової статистики, яку можна отримати зі статистичних даних в Інтернеті, можна обчислити ймовірність виграшу кожним гравцем гейму, потім сету і, нарешті, матчу. Саме завдяки припущенню незалежного однакового розподілу очок можна побудувати ймовірнісні моделі для етапів матчу – ланцюги Маркова.

Однак прогнозування шляхом ймовірнісних моделей не може охопити усі фактори та метрики, доступні для аналізу, незважаючи на різноманіття підходів побудови моделі. Беручи до уваги об'єм доступних статистичних даних тенісу, можна запропонувати альтернативний підхід до прогнозування тенісних матчів – машинне навчання. Параметри гравців матчу разом з результатом матчу можуть скласти навчальну вибірку. Алгоритм машинного навчання з учителем може використовувати цю вибірку для побудови функції передбачення результатів нових

матчів. Незважаючи на те, що машинне навчання, безумовно, оптимізує вирішення проблеми прогнозування тенісу, цей підхід до недавнього часу залучав значно менше уваги дослідників, ніж стохастичні ієрархічні методи. У більшості досліджень застосування машинного навчання до тенісу використовуються логістична регресія та нейронні мережі.

Найчастіше при побудові автоматизованих додатків використовують саме стохастичні моделі і надає користувачам ймовірності перемоги кожного гравця із супутньою статистикою, яку пропонується аналізувати самостійно.

Хоч прогнозування спортивних змагань й часто розглядається з точки зору букмекерства та ставок, проте комбінування двох вищеназваних підходів для прогнозування, а саме - стохастичного та машинного -, можуть давати нові підходи та результати, допомагаючи особам, задіяним у процесі підготовки до спортивних змагань, досягати кращих результатів.

**Мета дослідження.** Метою проекту є підвищення ефективності прогнозування результатів спортивних матчів за допомогою ланцюгів Маркова та нейронних мереж.

**Завдання дослідження.** Завдання дослідження:

- розробити формальну постановку задачі прогнозування спортивних змагань;
- розробити новий системний підхід до прогнозування результатів спортивних змагань;
- розробити методологію прогнозування результатів спортивних змагань, яка поєднує в собі метод прогнозування ланцюгами Маркова та нейронними мережами;
- модифікувати алгоритм прогнозування результатів спортивних змагань за допомогою ланцюгів Маркова для прогнозування спортивних змагань;
- впровадити методологію прогнозування спортивних змагань у вигляді програмного продукту, який може використовуватися для прогнозування.

**Об'єкти дослідження.** Головним об'єктом дослідження є процес прогнозування результатів спортивних змагань.

**Предмет дослідження.** Предметом дослідження є методи та моделі прогнозування результатів спортивних змагань.

# 1 ОГЛЯД ПІДХОДІВ ТА ОСНОВНІ ПОЛОЖЕННЯ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ СПОРТИВНИХ ЗМАГАНЬ

## 1.1 Задача прогнозування результатів спортивних змагань

Одним з методів визначення мети є прогнозування. У свою чергу прогнозування є однією з найважливіших функцій управління [24].

*Прогнозування* - це вид пізнавальної діяльності людини, що забезпечує формування прогнозів розвитку об'єкта на основі ймовірного, науково обгрунтованого аналізу динаміки його станів. Слід розрізняти такі види прогнозування: науковий, емпіричний, інтуїтивний, емоційний.

Для прогнозування широко використовуються наступні методи:

- метод логічного аналізу;
- метод експертних оцінок;
- метод екстраполяції;
- аналогія;
- інтуїтивне передбачення і гіпотеза;
- моделювання;
- систематизація;
- класифікація;
- періодизація.

Розробка прогнозів у спорті здійснюється у формі конкретизації передбачення перспектив розвитку того чи іншого процесу або явища, характерного для спортивної діяльності.

*Завдання прогнозування* зводиться до виявлення ймовірного розвитку того конкретного явища, яке відповідає знанню, відображає передові тенденції і визначає процес і результат заданого ефекту.

У спорті прогнозуванню піддаються різні процеси і явища: тенденції розвитку спорту та перспективи розвитку окремих видів спорту, системи змагань, спортивної підготовки, техніки і тактики окремих видів спорту, прогнози зростання спортивних рекордів, співвідношення сил на міжнародній і вітчизняній спортивній арені,

техніко-тактичних і функціональних можливостей окремих спортсменів (команд), прогнозування результатів окремих поєдинків тощо.

Наукове прогнозування в спорті [8] ґрунтується на використанні методу екстраполяції (передбачає поширення висновків, отриманих з спостереження під однією частиною явища, на інші його частини). Екстраполяція дозволяє здійснити прогнози зростання світових рекордів на основі вивчення відповідних закономірностей у попередні роки. Екстраполяцію доцільно використовувати в комплексі з методом моделювання та експертних оцінок.

Правила гри в теніс є ні що інше як аксіоми тенісу, що задають чіткі, певні параметри гри. Тенісистам і любителям цього виду спорту відома своєрідна арифметика тенісу - підрахунок очок, в якому присутній апарат математики. Різноманітність ударів, у напрямку, силі, дальності, висоті польоту, вирощений з тією чи іншою швидкістю навколо осей різної орієнтації тощо., все це сьогодні моделюється математичними об'єктами, які потім в сучасних умовах піддаються комп'ютерному дослідженню. Ніхто інший, як математик будує математичну модель гри в теніс, вивчення якої дозволяє відповісти на багато питань, що стосуються структури тенісного матчу. Вивчення моделі проводиться різним математичним інструментарієм, в тому числі ймовірісно-статистичними фактами, поняттями, законами.

Стан системи - гра в теніс - визначається рахунком в межах геймів [7]. При цьому перехід з одного стану в наступне (рахунок) залежить тільки від поточного стану і, звичайно, від ймовірності переходу, проте він не залежить від попередніх станів. Зауважимо, що будь-яка система, в якій перехід з одного стану в інший не залежить від передісторії процесу, а залежить тільки від поточного стану, називається в теорії ймовірностей марківським ланцюгом або ланцюгом Маркова (А.А. Марков, 1856-1922). Таку ланцюг потім вивчають ймовірісно-статистичними засобами і засобами теорії графів. Звичайно, будь-яка модель є наближеною.

В процесі гри кожен тенісист враховує свої помилки технічного і тактичного характеру, пристосовується до манери, темпу, стилю гри супротивника, іншими словами, навчається і вдосконалює свою гру.

## 1.2 Основні положення та елементи турнірів великого тенісу

Теніс - ракетковий спорт, в який можна грати або проти одного супротивника (одиначний розряд) або між двома командами з двох гравців (парний). Для простоти ми зупинимося лише на моделюванні одиночних ігор у тенісному матчі [9].

Мета гри — перекинути ракеткою м'яч через сітку на бік партнера так, щоб м'яч упав у межах майданчика суперника і той не зміг його відбити назад. Для цього потрібно відбити м'яч, за допомогою ракетки так, щоб він потрапив на половину поля суперника. Останній при цьому повинен промахнутися. Теніс цікавий тим, що він не обмежений за часом. Головне тут — підрахунок очок. Вони підраховуються протягом двох етапів: гри та партії. Починається гра з першої подачі. З неї ж починають розігрувати очко. Перша подача завжди здійснюється з першого «квадрата»: праворуч від центральної лінії. Завдання гравця — перекинути м'яч в протилежне поле суперника по діагоналі. Якщо м'яч потрапляє в сітку або за лінію подачі, то судді оголошують «аут», і гравець проводить другу подачу. Якщо і друга спроба не вдалася, то очко записується противнику. Коли м'яч перелетів сітку, але торкнувся її, тоді подача переграється. Після кожного розіграного очка, гравці міняють квадрати: переходять на протилежну сторону від центру. Зміна подачі м'яча відбувається через партію [52].

Розіграш кожного очка починається з подачі м'яча. Гравець, із задньої лінії перекидає м'яч на діагонально протилежно розташоване поле супротивника. Перша подача завжди здійснюється із правого боку від центральної лінії поля. Право вибору одного з боків корту, або черговість подачі в першому геймі, визначає жереб перед початком розминки. Гравець або пара, яка виграла жереб можуть обрати:

- подавати чи приймати в першому геймі матчу, тоді суперник повинен обрати на якому боці корту буде починати грати в першому геймі;
- бік корту, тоді суперник повинен обрати подачу або прийом у першому геймі;
- передати право вибору супернику.



Право подачі постійно протягом гри переходить від одного гравця до іншого.

Порушення, що зустрічаються при подачі:

- м'яч влучає за лінію площі подачі або в сітку - у цьому випадку гравець, що подає, має право на другу подачу;
- при повторному влученні в сітку чи за лінію площі подачі, очко присуджується стороні, що приймає - така ситуація називається *подвійна помилка*;
- заступ гравця, що подає, за задню лінію;
- м'яч торкається сітки — гравець повинен переграти подачу.

Період тенісного матчу, протягом якого подає один гравець, називається *геймом* [5]. Гейм починається з рахунку 0-0. Перші два розіграші м'яча дають супротивникам по 15 очок. Наступна подача приводить до рахунку 40, наступний розіграш приводить до виграшу гейму у випадку, якщо супротивник має рахунок 30 або менше. Якщо обоє гравці мають по 40, то виграш наступної подачі надає перевагу. Перемагає в геймі гравець, який має перевагу і виграв наступну подачу.

Розіграш *сету* йде доти, поки один із гравців не виграє 6 геймів. При рахунку 6-5 розігрується ще один гейм. Якщо рахунок стає 7-5, сет закінчується. При рахунку 6-6 розігрується тай-брейк.

Тенісний *матч* йде доти поки один із гравців не виграє два сету. Але на турнірах “Великого Шолому” у змаганнях чоловіків тенісний матч продовжується до трьох виграних сетів одним із суперників.

При рахунку в сеті 6 — 6 розігрується *тай-брейк*. Відповідно до черговості гравець робить першу подачу, потім супротивник робить дві подачі, далі зміна йде через дві подачі. Виграє тай-брейк той, хто перший набирає 7 очок із різницею в 2 очка. Тай-брейк проводиться доти, поки не буде досягнута різниця у два очки. Зміна сторін корту відбувається після кожних 6 очок.

Існують різні *ситуації м'яча* в грі. Якщо не зафіксована помилка та не повідомлено про перегравання, то м'яч знаходиться у грі з моменту введення його у гру гравцем, що подає та до закінчення розіграшу очка. Якщо м'яч торкається лінії, це вважається як торкання м'ячем частини корту, що обмежена цією лінією. Якщо

м'яч знаходиться у грі і торкається постійного обладнання після правильного потрапляння в корт, гравець, що завдав цього удару, виграє очко. Якщо м'яч, що перебуває в грі, торкається постійного обладнання до торкання землі, гравець, що завдав удару, програє очко.

Також є *специфіка порядку виконання подач*. По закінченні кожного гейму приймаючий стає той, хто подає, а той, хто подає стає приймаючим на весь наступний гейм. У парному матчі гравцям пари, яка повинна подавати, слід визначити того, хто подає в першому геймі. Перед другим геймом те ж саме повинні вирішити їхні суперники. Партнер гравця, що подав у першому геймі, повинен подавати в третьому, а партнер гравця, що подав у другому геймі, повинен подати в четвертому геймі. Ця послідовність зберігається до кінця сету.

При *подачі* безпосередньо перед рухом подачі гравець, що подає повинен стояти обома ногами за задньою лінією та всередині умовного продовження центральної мітки та бокової лінії. Він має підкинути м'яч рукою в будь-якому напрямку і вдарити по м'ячу ракеткою до того як той впаде на землю. Рух подачі вважається завершеним в момент нанесення удару по м'ячу, або в момент, коли гравець промахується ракеткою по м'ячу. Гравець, який володіє лише однією рукою, може використовувати ракетку для підкидання м'яча.

*Процедура подачі* в стандартному геймі наступна: гравець, що подає по чергово займає положення на різних половинах свого боку корту, починаючи з правої половини в кожному геймі. У тай-брейку подача повинна проводитися по чергово з кожної половини боку корту того гравця, що подає, починаючи з правої половини корту. М'яч після подачі повинен перелетіти через сітку та торкнутися частини квадрата подачі, який розташований по діагоналі, навпроти, перед тим як гравець, що приймає, відіб'є його.

Останній сет у грі здебільшого грається без тайбрейка, проте в фіналах Відкритого чемпіонату США тайбрейк розігрується навіть у останньому сеті. Система рахунку в геймах 15, 30, 40 пішла від *фр. quinze, фр. trente, фр. quarante*, що у французькій промові становить співзвучну послідовність [37].

Щоб перемогти в «геймі», гравцеві потрібно виграти чотири очка, причому перевагу над суперником має бути в два і більше очка. Гейм починається з нульового рахунку. Перші два виграних очка зараховуються як 15 і 30 відповідно. Третє вигране очко відзначається цифрою 40. Четверте — при перевазі над опонентом в два очка обіцяє виграш «гейми». Може бути рахунок 30:30 або 40:40 — він називається «рівно». При такому рахунку тенісистові для перемоги необхідно виграти два очка поспіль.

Виграш в шести «геймах» приносить перемогу в «сеті». За умови, що перевага над суперником — дві або більше ігор. При рівному рахунку розігрується додатковий «гейм» — «тай-брейк». Завдання гравця — набрати 7 очок. Тоді перемога в «геймі» означає виграш сету. Ну, а щоб перемогти противника, тенісист зобов'язаний виграти два сету (якщо гра складається з трьох) або три (гра в п'ять партій) [56].

Корт (поле для гри) [13] ділиться на чотири зони: перша пара розташовується впоперек (за рахунок сітки), а друга вздовж неї (по центру). Таким чином, виходить 4 квадрата. Дві половини корту розділяє сітка висотою 3 фути в центрі та 3 фути 6 дюймів біля стовпчиків. Сітка повинна бути розташована таким чином, щоб повністю закривала простір між двох стовпів і мала розмір клітин такої величини, щоб м'яч не міг пролетіти крізь них. У центрі сітка повинна бути обмежена ремнем на висоті 3 футів (0,914 метра). Металевий кабель і верх сітки повинні бути закриті смугою із тканини. Ремінь і смуга тканини повинні бути повністю білих кольорів.

Тенісний корт може мати різне покриття. Корти розрізняються за покриттям. Класичними покриттями вважаються трава та ґрунт, проте в останні роки виникло дуже багато різних синтетичних видів покриття, які вимагають менше догляду, забезпечуючи добру якість поверхні. В залежності від покриття корти поділяються на швидкі і повільні в залежності від відскоку м'яча. Найшвидшим покриттям із тих, що використовуються в турнірній практиці є трава. Найповільнішим — ґрунт. Швидкість корту дає перевагу гравцям із певним стилем, тож кожен тенісист має свої улюблені покриття. Від покриття залежить стиль, тактика, стратегія та видовищність гри.

При всьому різноманітні існуючих на сьогодні покриттів, їх умовно поділяють на чотири види:

- трава (натуральна) — умовно швидке покриття, має низький та непередбачуваний відскік;
- ґрунт — повільне покриття;
- хард — швидке покриття;
- килим — покриття для залів, може бути як швидким так і повільним.

### 1.3 Додаткові правила, техніка та тактика гри, традиції екіпіровки

Правила тенісу включають у себе наступні *додаткові правила*:

- лінія обводу ігрового поля вважається частиною поля;
- м'яч, що зачепив сітку й упав на бік супротивника, в ігровий відрізок зараховується як очко (але не під час подачі);
- прийом подачі можливий тільки після відскоку м'яча, у той час, як під час гри м'яч може бути відбитий і до торкання ним поверхні корту;
- не зараховується очко у випадках торкання м'яча тіла спортсмена, у випадках прийому м'яча до перетинання ним лінії сітки, при торканні гравцем сітки або стійки сітки ракеткою, рукою або іншою частиною тіла.

В тенісі є одиночні та парні ігри. Парна гра проводиться на розширеному майданчику. Парна гра, в якій в одній парі грають чоловік та жінка, називається мікстом. На любительському рівні або ж на тренуваннях практикують також гру у трьох — по чергово один проти двох. Одиночний спортсмен змагається за звичайними правилами для одиночних ігор, двоє його суперників — за правилами для парних ігор. Якщо гейм виграє одиночний тенісист, йому зараховують два очки, якщо виграє пара тенісистів, то кожен із неї здобуває одне очко. Після кожного гейму одиночний тенісист міняється місцем з тим із пари, що грав справа. Рахунок ведуть так само, як і у звичайній грі. Із трьох учасників виграє сет той, що набере принаймні шість очок, маючи не менш як двухочковий відрив від кожного з двох суперників.

Основні *компоненти тенісної техніки*: біг, хват ракетки та удар по м'ячу. Надзвичайно важливо також спостереження за м'ячем і суперником, а також вибір правильної позиції на корті [28].

Основний елемент — *тенісний крок*, тобто спосіб переміщення тенісиста по корту. У сучасному тенісі розрізняють: звичайний, перехресний (або околишній) і приставний кроки. Вибір кроку визначається ігровою ситуацією та місцем знаходження тенісиста на корті.

При виконанні різних технічних прийомів багато значить *хватка*, тобто те, як спортсмен тримає ракетку. Розрізняють три хватки: східна, західна та континентальна. У кожній з них є певні переваги та недоліки, а основне розходження між хватками полягає в положенні долоні тенісиста відносно рукоятки ракетки. При західній хватці долоня перебуває під рукояткою, при континентальній (її також називають універсальною) — над рукояткою, а при східній — за нею. У сучасному тенісі найпоширеніша східна хватка, саме з неї починають освоювати ази техніки новачки. Але при будь-якому типі хватка повинна бути впевненою та міцною.

*Удари по м'ячу* розрізняються залежно від техніки їхнього виконання, траєкторії м'яча до та після зіткнення з ракеткою та інших моментів. В арсеналі сучасних тенісистів чимало різноманітних ударів: *смейш* (потужний удар над головою по м'ячу, що опускається), скорочений удар (після якого м'яч приземляється відразу за сіткою на «чужій» половині площадки), *топспін* (кручений удар з високою траєкторією польоту м'яча), *удар у протихід* (тобто в те місце корту, яке суперник щойно залишив) та інші. У кожного із гравців є свої коронні удари. Так, наприклад, швед Бйорн Борг був визнаним майстром сильних кручених ударів із задньої лінії [28].

Якість гри тенісиста залежить не тільки від фізичної сили. Австралієць Род Лейвер, який не відзначався зовнішніми даними, відрізнявся, проте, неймовірною витривалістю та мав воістину залізну п'ясть, що дозволяло йому бити хльостко та потужно. Нерідко фахівці використовують таке поняття, як почуття м'яча, тобто навички володіння м'ячем: уроджені й придбані завдяки тренуванням й ігровій практиці. Крім того, різноманіття чинених тенісистом рухів (біг, нахили, стрибки, удари тощо) вимагає від нього різнобічної фізичної підготовки.

Тенісист прагне нав'язати суперникові свою гру, що дозволила б йому максимально використати власні сильні сторони та нейтралізувати переваги суперника. Також дуже важливим є не тільки улюблені технічні прийоми та «місце»

гри, але і її темп (фахівці виділяють швидкий, повільний і рваний темпи гри). Приміром, зміна ритму в сполученні зі зміною довжини ударів вважається найвигіднішою при грі із задньої лінії [28].

Залежно від розряду змагань виділяють тактики одиночної та парної гри, кожна з яких, у свою чергу, ділиться на тактику того, хто подає (або пари, що подає) і тактику приймаючого (приймаючої пари).

У тенісі, як і у всякій грі, є свої неписані закони та заповіді. Наприклад, одна з них говорить, що «кручені» м'ячі частіше приземляються в межах майданчику. При проробці плану гри на конкретний матч визначаються і з покриттям майданчику («швидкі» та «повільні» корти), а також знанням сильних і слабких сторін суперника і його фізичних кондицій: так, зазвичай тенісисти намагаються не допустити виходу суперника до сітки, але у випадку, якщо він погано грає з льоту, можуть навмисно «викликати» до неї. Загальний успіх у грі залежить не тільки від якості виконання награних комбінацій і прийомів, але і від уміння в потрібний момент зімпровізувати, зіграти зненацька для супротивника. Тенісний матч — це свого роду «забіг на довгу дистанцію», у якому є кілька кульмінаційних моментів. Уважається, що в геймі ключовим є розіграш третього очка, а в партії — четвертий гейм. Варто також пам'ятати, що гра — це не тільки єдиноборство технік, але й боротьба нервів.

Тактика парної гри помітно відрізняється від тактики тенісистів-«одинаків». Досить сказати, що якщо в одиночній грі очки в основному виграються від задньої лінії, то в парній — біля сітки, при цьому парні зустрічі відрізняються більш високим темпом гри. Проте і тут подача є важливим тактичним елементом. За статистикою, у парній грі більшість геймів виграється саме на своїх подачах. Гра в парі вимагає від спортсменів повного взаєморозуміння та виняткової концентрації, адже тут їм доводиться тримати в полі зору не лише м'яч, але і партнера та обох суперників. Зазвичай гравці в парі підбираються так, аби вони максимально доповнювали один одного — як за технікою, так і за темпераментами. Індивідуальними особливостями гравців визначають їхнє первісне розташування на майданчику, а також порядок подач: сильніший у цьому компоненті гри партнер у кожній партії подає першим. Крім загальних, у парній грі необхідно дотримуватись і деяких специфічних

неписаних законів: наприклад, «свічу» завжди відбиває той з партнерів, через голову якого вона пущена. Є тут і свої психологічні нюанси. Для граючих у парі дуже важливе вміння адекватне зреагувати на промах партнера, морально підтримати його в напружений момент гри. За свідченням очевидців, природженим парником був француз Жак Брюньон, легенда довоєнного тенісу [18].

*Екіпіровка* тенісиста включає спеціальний спортивний одяг і взуття. Загально прийнятий колір тенісної форми — білий. Він є практичним: білий колір завдяки своїм властивостям добре «відбиває» тепло, що дуже важливо при грі у спеку. У той же час біла форма є даниною історичним традиціям: стіни у французьких «домах для гри у м'яч» були чорного кольору, і гравці, одягнені у білу форму, найкраще помітні на такому фоні. Екіпіровка тенісиста з часом зазнала суттєвих змін. Так, на заміну фланелевим штанам тенісистів-чоловіків прийшли шорти, а колись довгі спідниці тенісисток стали значно коротшими та зручнішими.

Взуття тенісиста повинно бути легким та міцним. Найважливіша частина — підошва, характеристики якої відрізняються в залежності від покриття корту. Так, рифлена підошва, що добре запобігає ковзанню ніг на ґрунтових кортах, на жорстких кортах може призвести до травми через надлишкове тертя. Для кортів з асфальтовим, дерев'яним та іншими видами жорсткого покриття більше підходить гладка підошва [59].

Тенісну форму можуть доповнювати кепки від сонця (або бандани) та напульсники, які запобігають розтягненню м'язів рук і потінню долонь.

## **1.4 Тенісні турніри**

Теніс протягом тривалого часу був переважно розповсюджений в англomовних країнах — Австралії, Великій Британії, США. Він також був популярний у Франції. Чотири найбільших турніри — Вімблдонський турнір, Відкритий чемпіонат США, Відкритий чемпіонат Франції та Відкритий чемпіонат Австралії (проводиться з 1905) стали найпрестижнішими в тенісі.

У 1933 австралієць Джек Крофорд (англ. Jack Crawford) виграв міжнародні аматорські ігри в Австралії, Франції (Роланд Гаррос), Великій Британії (Вімблдон) і

досяг фіналу Міжнародного аматорського чемпіонату США у Форест-Гілл. Саме тоді, напередодні останнього з турнірів, обговорюючи шанси Крофорда на перемогу в усіх чотирох найвизначніших змаганнях року, журналісти Джон Кієран і Аллісон Дензіг уперше вжили термін «Великий шолом» (англ. Grand Slam) (за аналогією із бриджем і гольфом). Перелічені чемпіонати почали набирати значимість у тенісному світі у зв'язку з тим, що збірні цих чотирьох країн були єдиними, хто здобував перемоги у “Кубку Девіса”, який на той час був набагато престижнішим в аматорському тенісі, ніж усі інші турніри. Виграти «Великий Шолом», тобто перемогти у всіх чотирьох цих турнірах протягом одного сезону — є вищою метою для тенісистів [38].

Найпрестижнішими в тенісі є чотири турніри великого шолому:

- відкритий чемпіонат Австралії;
- відкритий чемпіонат Франції;
- Вімблдонський турнір;
- відкритий чемпіонат США.

Найпрестижнішими командними змаганнями є:

- кубок Девіса — змагання збірних країн серед чоловіків;
- кубок Федерації — змагання збірних країн серед жінок;
- кубок Хопмана — змагання змішаних чоловічо-жіночих збірних.

Менш престижний Олімпійський тенісний турнір, який відбувається кожні чотири роки під час літніх Олімпіад.

Окрім того, професійні тенісні організації організують серії турнірів різної сили, так звані тури, охоплюючи ними всіх гравців організацій, тенісну молодь і ветеранів.

Тенісні турніри здебільшого проводяться за олімпійською системою. Пари визначаються шляхом жеребкування з попереднім розсіюванням. Зрідка проводяться колові турніри, або турніри, в яких спочатку граються матчі в групах з подальшим переходом на олімпійську систему [66].



## 1.5 Дані для аналізу

Статистичні дані тенісних матчів широко доступні в Інтернеті. Офіційні сайти турнірів, наприклад, ATPworld [21], надають інформацію про гравців та результати матчів, а також результативність спортсмена за кожного матчу. Деякі джерела, наприклад, Tennis-Data [22], надають статистичні дані в структурованій формі (CSV-або Excel-файли). Доступні платні бази даних – більш комплексні, на більш довгі періоди і з кращою точністю, наприклад, база OnCourt [25].

Для повного розуміння метрик і факторів необхідно розбиратися у термінології тенісу. Основна термінологія тенісу являється наступною:

- *бекхенд або удар закритою ракеткою* — удар з незручного боку, зліва для правші, справа для шульги;
- *брейк* — виграш подачі супротивника;
- *брейкпойнт* — розіграш, у якому один із супротивників може здобути брейк;
- *геймпойнт* — розіграш, у якому один із супротивників може виграти гейм на своїй подачі;
- *волей або удар зльоту* — удар по м'ячу перш, ніж той встиг опуститися на корт;
- *дропшот або укорочений удар* — несподіване підрізання м'яча таким чином, щоб він ледь-ледь перелетів через сітку й опустився на корт майже без відскоку;
- *ейс* — подача навиліт, при якій супротивник не зміг дотягнутися ракеткою до м'яча;
- *епроуч* — повільний удар із сильною підрізкою з метою виходу до сітки;
- *крос* — діагональний удар;
- *лоб або свічка* — високий удар з метою перекинути супротивника, який вийшов до сітки;
- *матчпойнт* — розіграш, у якому один із супротивників може виграти гейм, сет і матч;

- *напівволей або удар із напівльоту* — удар по м'ячу, який щойно відбився від землі й підіймається вгору;
- *реверс* — удар оберненою (закритою) стороною ракетки високо над головою;
- *сетпойнт* — розіграш, в якому один із супротивників може виграти сет;
- *смейш* — сильний удар піднятою над головою ракеткою;
- *слайс або підрізка* — удар, при якому м'ячу надається обертання в напрямку протилежному до напрямку руху - завдяки ефекту Магнуса м'яч стелиться над кортом, продовжуючи свою траєкторію; застосовується для гри із задньої лінії;
- *топ-спін* — удар, при якому м'ячу надається сильне обертання в напрямку руху; характеризується високим відскоком і траєкторією, яка закручується донизу завдяки ефекту Магнуса; застосовується при грі із задньої лінії;
- *тай-брейк* — спеціальний гейм для швидкого визначення переможця рівного сету;
- *форхенд або удар відкритою ракеткою* — удар зі зручного боку, справа для правші, зліва для шульги.

Найбільш релевантні дані, які можна взяти з подібних баз даних, що представлені в таблиці 1.1.

Таблиця 1.1 – Дані, за допомогою яких можна аналізувати тенісні матчі

<i>Дані про гравця</i>	Ім'я
	Дата народження
	Країна
	Призовий фонд
	Рейтинг за очками
	Загальний рейтинг АТР або WTA
<i>Дані про матчі</i>	Назва турніру
	Тип турніру (наприклад, Великий шолом)
	Покриття корту
	Місце розташування (країна, координати)
	Дата
	Результат (рахунок по сетах)
	Призовий фонд
	Коефіцієнти (від Pinnacle)
<i>Поматчева статистика для обох гравців</i>	Відсоток виграшу на першій подачі
	Эйсы
	Подвійні помилки
	Невимушені помилки
	Відсоток очок, виграних при першій подачі
	Відсоток очок, виграних при другій подачі
	Відсоток очок, виграних при прийомі

	Переможці
	Брейк-пойнти (виграні, всього)
	Виходи до сітки (виграні, всього)
	Всього виграних очок
	Найшвидша подача
	Середня швидкість першої подачі
	Середня швидкість другої подачі
	Коефіцієнти (від Pinnacle [57])

Для моделювання матчу можуть бути важливі і такі дані як статистика по рахунках і по очках для кожного гравця. Ці дані можна отримати шляхом парсингу сайтів. Важливо відзначити, що за допомогою технології відстеження м'яча HawkEye [29] для багатьох турнірів можна отримати дані більш високої якості і деталізації, наприклад, положення м'яча і гравця в будь-який момент матчу. Однак асоціація АТР, що володіє цими даними, не видає ліцензії на їх використання третім сторонам [21].

## 1.6 Збір вхідних даних

Наявність великого масиву вихідних даних для навчання - один з головних чинників, що визначають можливість успішного вирішення поставленого завдання. Але оскільки прогнозіст планує передбачати результати матчів в різних видах спорту, в різних за рівнем змаганнях, він повинен це враховувати вже на етапі підбору навчальних даних і, при необхідності, формувати кілька різних навчальних вибірок [15].

Ще один важливий фактор - якісний склад даних в навчальних векторах. Чим більше різних факторів буде враховано, тим глибше можна виконати кластеризацію для отримання нових даних для гравця в процесі прогнозування. Ось короткий перелік груп чинників, які можна враховувати:

- фактор поля (своє, чуже, нейтральне);
- покриття поля;
- оперативна інформація: сила гравців на поточний момент і динаміка їх зміни (власне чисельне значення сили, отримане на основі будь-якого рейтингу, який оновлюється регулярно, різниця в силі, співвідношення сил);
- статистика матчів.

В яких саме кількісних характеристиках описувати наведені групи чинників – вибір дослідника.

Окремого обговорення вимагають чинники, які важко оцінити кількісно: травма або дискваліфікація провідного гравця, зміна тренера, турнірна ситуація. Теоретично враховувати їх, хоч і складно, але можна. Однак на практиці найскладнішим виявиться зібрати і представити їх в навчальній вибірці. Саме з цієї причини є сенс враховувати їх уже після отриманого рішення на заключній стадії, коректуючи ймовірності [23].

## **1.7 Кластеризація при передбаченні результатів змагань**

Кластеризація [19] - це завдання розподілу великого масиву наявних даних по групах (кластерах) на основі загальних або, в разі їх кількісного виміру, близьких за величиною ознак. Таким чином, окремо взяті вектори даних (приклади), що сформували в результаті роботи штучної нейронної мережі кластер, по-перше, повинні бути дійсно схожі між собою, а по-друге, самі і визначають властивості утвореного ними кластера. Процедура опису якісних і кількісних властивостей кластерів називається кластерним аналізом. Отримане опис цінний сам по собі, але також може бути використаний для подальшої роботи з уже навченою нейронною мережею, коли виконується віднесення нового прикладу до одного з уже існуючих кластерів. Якщо процес налаштування мережі і процедура кластерного аналізу були виконані якісно, на досить великому обсязі коректних навчальних даних можна з дуже високою часткою впевненості стверджувати, що опис відповідного кластера буде відповідати і цьому новому наприклад.

Вирішуючи задачу прогнозування результату матчу, потрібно знати, що її відповіддю ніколи не повинно бути передбачення єдиного результату, так як на нього можуть впливати несподівані фактори. Питання лише в частоті реалізації тієї чи іншої події. Таким чином, в результаті рішення задачі повинні бути отримані ймовірності кожного з можливих результатів гри, і їх сума повинна бути строго дорівнює одиниці [19].

Серед методів нейромережевої кластеризації найчастіше використовують три наступних:

- нейронні мережі Кохонена;
- нейронні мережі адаптивного резонансу ART-2;
- нейронні мережі на основі радіально-симетричних (радіально-базисних) функцій.

Перераховані архітектури нейронних мереж навчаються самостійно на спеціально складених для цього вибірках даних. Структура їх може бути фіксованою, але кращий варіант самоорганізації, коли в процесі навчання не тільки налаштовуються вагові коефіцієнти, але й формується сама структура мережі: утворюються нові кластери. Якщо кількісні і якісні характеристики прикладу не підходять досить близько до жодного з наявних кластерів, а також можливе видалення кластера (нейрона), якщо протягом поточної доби навчання він не був задіяний. Відмінності зазначених архітектур полягають в основному лише в використовуваному алгоритмі корекції ваг, про що можна дізнатися зі статей, присвячених реалізації відповідних методів [19].

Рішення ж конкретного завдання прогнозування результату матчу можна представити у вигляді послідовності наступних етапів:

- збір вихідних даних про які мали місце спортивних подіях того ж виду спорту і рангу;
- вибір архітектури нейронної мережі для вирішення задачі;
- вибір або розробка програмного продукту для розв'язання задачі;
- формування навчальної вибірки і її структуризація під вимоги використовуваного програмного забезпечення;

- налаштування алгоритму навчання;
- навчання та кластерний аналіз;
- практичне використання навченої нейронної мережі для прогнозування вірогідності можливих результатів.

## 1.8 Відомі результати аналізу факторів результативності гравців

Прогнозування - наукова основа планування і управління спортивною діяльністю. Прогноз необхідний ще до прийняття відповідних планових рішень. Спортивний прогноз не є остаточною рекомендацією або вибором, це тільки багатоваріантна, науково обґрунтована об'єктивна оцінка. З усіх можливих варіантів на основі критерію оптимальності розвитку спортсмена вибирається і затверджується найбільш ефективний план тренувань гравця.

Основаючись на роботі Ма, Ліу і Тана [60] можна говорити про наступні результати у дослідженнях в області надання рекомендацій щодо покращення результативності тенісистів в рамках "Великого Шолома".

Шанси виграти матч у чоловічих синглах "Великого Шолома" можуть бути покращені завдяки підвищенню показників перших подач, ейсів, виграшу очок других подач, першого прийому, другого прийому, а також переведених та збережених брейк-поінтів. Чоловіки-гравці зі зростом менше 180 см мали менше шансів виграти матч, ніж ті, що мають зріст від 181 см до 185 см, але не менш ймовірно, ніж ті, що вище 186 см. Тим не менше, це дослідження показує, що позитивна кореляція між зростом та перемогою не є лінійною, а зріст не дає позитивного впливу на результати, якщо гравці перевищують зріст 186 см. Крім того, відмінність між зростом гравця та його супротивником у конкретному матчі не є значним фактором впливу на результати матчу. Цей висновок дає важливі рекомендації щодо вибору молодих тенісистів чоловічої статі, оскільки їхній остаточний зріст може вплинути на майбутні виступи в матчах "Великого Шолома".

Кількість професійних років не впливає на результати матчів. Є два можливі пояснення щодо впливу професійних років. По-перше, у професійних гравців занадто багато непередбачуваних переривань у кар'єрі, таких як травми, шлюб тощо. По-

друге, показники найуспішніших гравців на такому високому рівні не можуть суттєво покращуватися з часом та досвідом. На відміну від цього, бадьорість та енергетична перевага молодих гравців, як виявляється, перевершують перевагу досвіду, яку мають старші гравці. Ймовірність виграшу гравця з вищим рейтингом зменшується, оскільки цей гравець конкурує з молодшими гравцями.

Відкриттям дослідження стало те, що домінуюча рука не була основним фактором перемог матчів: гравець-шульга не має переваги, коли конкурує з гравцями, у яких основною є права рука. Як запропонував Лофінг [47], більшість гравців - шульги, тому їх загальна перевага зменшена, отже у праворуких гравців є більше можливостей виграти. Виходячи з цього, дослідження стверджує, що перевага лівої руки зникає в міжнародних турнірах, тому що стандартні висококласні гравці навчаються таким чином, щоб пристосуватися до відбивання м'яча та куту бічного льотного м'яча від гравців, що являються шульгами. Тим не менш, Коррал і Прієто-Родрігес [39] заявляли про свою перевагу в тому, що були шульгою, коли нижчий по рівню гравець змагається з праворуким, вищим по рейтингу суперником на турнірі "Великого Шолома". Однією з можливих причин є те, що гравці, що стоять вище за рейтингом, не звикли до стилю гри нижчого по рівню суперника.

Відповідно до теорії досягнення цілей, конкретні цілі, поставлені гравцю підвищують ефективність за допомогою зменшення уваги, розстановку зусиль, наполегливості та розвиток когнітивних стратегій, що наведено в роботі Лока та Латана [46]. Тому тренери з тенісу можуть використовувати результати статистичних даних турнірів "Великого Шолому" для встановлення конкретних цілей для своїх гравців: це допомагає підвищити ефективність тренувань. Тренери можуть записувати результати та показники своїх гравців щоденно під час тренування та порівнювати їх з результатами переможців. Таким чином, вони можуть точно знати, наскільки їхні навички та продуктивність відрізняються від найуспішніших гравців, а також ставити конкретні цілі та плани навчання.



## **1.9 Відомі результати дослідження прогнозування результатів спортивних змагань**

Сучасні моделі прогнозування тенісу можуть бути засновані на ієрархічних стохастичних виразах. Кноттенбельт [68] уточнив моделі Барнета [63], використавши для обчислення ймовірності виграшу очка при подачі тільки матчі зі спільними суперниками гравців, замість всіх минулих суперників. Цей підхід дозволяє знизити похибка, що виникає через те, що гравці в минулому зустрічалися з суперниками різного рівня.

Мадурська [32] далі розширила модель загального суперника Кноттенбельта [68], використавши різні ймовірності виграшу очка при подачі для різних сетів. Таким чином, автор відмовилася від допущення неналежного однакового розподілу і її модель відображає накопичення фізичної втоми у гравця під час матчу.

Більшість опублікованих моделей машинного навчання використовують логістичну регресію. Кларк і Дайт [61] застосовують модель логістичної регресії до різниці за очками двох гравців в рейтингу АТР для прогнозування результату сету.

Сомбунокопан [33] навчив тришарову нейронну мережу прогнозувати результати тенісних матчів з використанням алгоритму зворотного поширення. Автор досліджував і порівнював кілька різних мереж з різними наборами вхідних параметрів. Краща нейронна мережа складається з 27 вхідних вузлів, що представляють такі ознаки матчу і гравця як поверхня корту, відсоток виграшу на першій подачі, другий подачі, відповідної подачі, брейк-пойнтах тощо. Автор стверджує, що точність прогнозу складає близько 75% при прогнозуванні результатів матчів турнірів Великого шолома в 2007 і 2008 років.

Сіпко [50] використовує логістичну регресію і нейронні мережі, перевіряючи моделі на вибірці з 6135 матчів турнірів АТР 2013-2014 рр., ROI найбільш точної моделі склав 4,35%, що за заявою учасника на 75% краще сучасних стохастичних моделей.

Фернандес [48] розробив підхід прогнозування, що включає в себе систему голосування, що складається з нечіткого класифікатора, нейронної мережі та

зваженого рівняння, за допомогою яких реалізується процес прогнозування результатів спортивних змагань.

### **1.10 Висновок до розділу**

Резюмуючи вищесказане можна зробити висновок, що незважаючи на існування великої кількості спроб точного прогнозування спортивних турнірів, область все ще вимагає дослідження та розробки алгоритмів прогнозування турнірів.

Підвищення ефективності

Метою дослідження є підвищення ефективності прогнозування результатів спортивних матчів.

Для досягнення мети необхідно виконати наступні задачі:

- розробити формальну постановку задачі прогнозування спортивних змагань;
- розробити новий системний підхід до прогнозування результатів спортивних змагань;
- розробити методологію прогнозування результатів спортивних змагань, яка поєднує в собі метод прогнозування ланцюгами Маркова та нейронними мережами;
- модифікувати алгоритм прогнозування результатів спортивних змагань за допомогою ланцюгів Маркова для прогнозування спортивних змагань;
- впровадити методологію прогнозування спортивних змагань у вигляді програмного продукту, який може використовуватися для прогнозування.

## 2 ІСНУЮЧІ ПІДХОДИ ДО ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ СПОРТИВНИХ ЗМАГАНЬ

### 2.1 Марківські ланцюги для прогнозування результатів тенісних турнірів

У математичному моделюванні існують як детерміновані, так і стохастичні моделі. І якщо з детермінованими моделями досить успішно працюють, то стохастичні моделі в значно більшій мірі є добре розробленим розділом математики, який у меншій мірі знаходить своє застосування на практиці.

В даний час все більше застосування може знаходити такий статистичний розділ математики, як теорія ланцюгів Маркова [14]. Цей розділ є частиною теорії ймовірності і тому придатний для стохастичного моделювання.

Стохастична модель (ймовірнісна модель) - модель, у якій використовується одна або більше випадкових величин для врахування невизначеності процесу, або в якій вхідні дані будуть представлені відповідно до деякого статистичного розподілу. Стохастична модель - математична модель, в якій параметри, умови функціонування і характеристики стану об'єкта, що моделюється представлені випадковими величинами і пов'язані стохастичними (випадковими, нерегулярними) залежностями, або вихідна інформація також представлена випадковими величинами. Отже, характеристики стану в моделі визначаються не однозначно, а через закони розподілу їх ймовірностей. Моделюються, наприклад, стохастичні процеси в теорії масового обслуговування, в мережевому плануванні та управлінні і в інших областях. При побудові С. м. застосовуються методи кореляційного і регресійного аналізів, інші статистичні методи.

Ланцюг Маркова в математиці [14] - це випадковий процес, що задовольняє властивість Маркова і який приймає скінченну чи зліченну кількість значень (станів). Існують ланцюги Маркова як з дискретним так і з неперервним часом.

В теорії імовірнісного моделювання до найбільш досліджуваних і досліджених відносяться моделі, у яких випадковий процес функціонування відноситься до класу марківських процесів, тобто, марківські моделі. Випадковий процес, що протікає в

системі, називається марківським, якщо для будь-якого моменту часу імовірнісні характеристики процесу в майбутньому залежить від його стану в даний момент і не залежить від того, коли і як система прийшла в цей стан.

Метод прогнозування на основі теорії ланцюгів Маркова може бути використаний для прогнозу безлічі показників, які змінюються з року в рік одночасно, але безпосередньо функціональні зв'язки між ними не встановлено через відсутність інформації або складності цих зв'язків.

Поширеним є метод визначення ймовірності виконання подій певного класу. Зв'язки між подіями такі, що якщо відома інформація про їх стан в попередній момент, то аналогічна інформація про більш ранні стани нічого нового не несе. Саме такі зв'язки між подіями і були досліджені російським математиком А.А. Марковим.

Поширеним способом візуального задання ланцюга Маркова є граф переходів. Вершини цього графа ототожнюються зі станами ланцюга Маркова, а орієнтовне ребро проходить з вершини  $i$  у вершину  $j$  лише у випадку, коли ймовірність переходу між відповідними станами нерівна нулю. Дана ймовірність переходу також позначається біля відповідного ребра.

Клаассен і Магнус [41] оскаржили теорію незалежного однакового розподілу, показавши, що очки в тенісі розподіляються незалежно і не однаково. Однак вони також показали, що відхилення від незалежного однакового розподілу настільки малі, що використання цього припущення часто дає коректні середні значення.

Цей факт дозволяє припустити, що для кожного отриманого очка у матчі результат не залежить від попередніх очок. Припустимо далі, що ми знаємо ймовірність виграшу очка при подачі кожним гравцем. Нехай  $p$  – ймовірність того, що гравець  $A$  виграє очко при подачі,  $q$  — ймовірність того, що гравець  $B$  виграє очко при своїй подачі. Використовуючи допущення незалежного однакового розподілу та ймовірності виграшу очок, можна побудувати марківську модель, що описує ймовірність перемоги гравця в геймі.

Виходячи з ідеї моделювання тенісних матчів з ланцюгами Маркова, Барнетт і Кларк [63] та О'Малі [44] розробили ієрархічні вирази для ймовірності того, що конкретний гравець виграє весь тенісний матч. Барнетт і Кларк [63] виражають

ймовірність того, що гравець  $A$  виграв гру при своїй подачі  $P_{\text{гейм}}$ , використовуючи наступне рекурсивне визначення:

$$P_{\text{гейм}}(x, y) = p \cdot P_{\text{гейм}}(x + 1, y) + (1 - p)P_{\text{гейм}}(x, y + 1) \quad (1)$$

Граничні значення записуються як:

$$P_{\text{гейм}}(x, y) = 1, \text{ якщо } x = 4, x - y \geq 2 \quad (2)$$

$$P_{\text{гейм}}(x, y) = 0, \text{ якщо } y = 4, y - x \geq 2 \quad (3)$$

$$P_{\text{гейм}}(x, y) = \frac{p^2}{p^2 + (1-p)^2}, \text{ якщо } x = 4, x - y \geq 2 \quad (4)$$

У формулах (1)-(4) -  $p$  - ймовірність того, що гравець  $A$  виграє очко при подачі, а  $x$  та  $y$  представляють кількість очок, виграних гравцями  $A$  та  $B$ , відповідно. Цей вираз явно відповідає ланцюжку Маркова.

## 2.2 Моделювання тенісного гейму

В даній роботі, з метою підвищення точності прогнозу, наведено варіант модифікації методу ланцюга Маркова.

Враховуючи те, що на даний момент існує вільний доступ до різноманітних статистичних даних тенісистів, то на початку матчу в якості ймовірності виграшу можемо обрати відповідні дані з попередніх матчів або тренувань гравця (відсоток подачі (прийому), виграшів з подачі (прийому), тощо). Згідно з цим можна побудувати 2 дерева – для подачі та прийому гравцем м'яча. Окрім того, постійне оновлення статистичних даних дає змогу у кожному наступному геймі враховувати результати попереднього. Також слід зазначити, що таким чином можна враховувати статистику ейсів, відсотки попадання 1-ї та 2-ї подач та подвійних помилок, що робить схеми Маркова максимально докладними.

Візьмемо за основу гейм і вважатимемо рахунок простором станів. Побудуємо модель Маркова для ланцюга гри. З ймовірністю  $p_A$  стан змінюється з  $a, b$  на  $a + 1, b$  і

з ймовірністю  $q_A = 1 - p_A$  він змінюється з  $a, b$  на  $a, b + 1$ . Таким чином, якщо  $P_A^{pg}(a, b)$  - це ймовірність того, що гравець А виграє гру, коли рахунок  $(a, b)$  ми маємо наступну формулу при обмеженнях. Барнетт і Кларк [63] виражають:

$$P_A^{pg}(a, b) = p_A P_A(a + 1, b) + q_A P_A(a, b + 1)$$

$$P_A^{pg}(a, b) = 1, \text{ якщо } a = 4, b \leq 2$$

$$P_A^{pg}(a, b) = 0, \text{ якщо } b = 4, a \leq 2$$

Слід також розглядати ситуацію рівного рахунку (наприклад, 30-30). Формула отримується з урахуванням того, що шанс виграти при рівному рахунку відбувається у вигляді геометричного ряду:

$$P_A^{pg}(3,3) = p_A^2 + p_A^2 2p_A q_A + p_A^2 (2p_A q_A)^2 + p_A^2 (2p_A q_A)^3 + \dots$$

Врахувавши наші обмеження, загальна формула для ситуації рівного рахунку:

$$P_A^{pg}(3,3) = p_A^2 / (p_A^2 + q_A^2)$$

Загальну візуалізацію ланцюга Маркова для тенісного гейму подано на рис. 2.1.

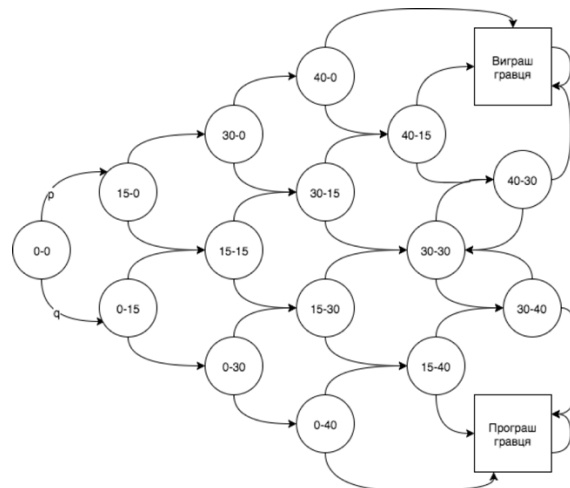


Рисунок 2.1 - Марківська модель гейму в тенісному матчі

### 2.3 Моделювання тенісного тай-брейку

Тайбрейк (від англ. tie — нічия і англ. break — ламати) — процедура швидкого визначення переможця в спортивних змаганнях, які закінчилися нічийним результатом у рамках, визначених основними правилами. Найвідоміші правила тайбрейку – тенісні.

Тайбрейк грається при рахунку 6:6 за геймами в сеті. При тайбреку кожен із гравців подає дві подачі в праву та ліву частини корту, після чого право подачі передається супротивнику. Зміна половинами корту відбувається після кожних шести подач. Ведеться підрахунок виграних м'ячів, і для перемоги потрібно виграти 7 м'ячів, але принаймні на два більше, ніж у супротивника.

Барнетт і Кларк [63] виражають згідно з попередніми міркуваннями про гейм тенісного матчу відповідні формули. Рекурентні формули для тайбрейку гейму матимуть вигляд:

$$P_A^{pgT}(a, b) = p_A P_B^{pgT}(a + 1, b) + q_A P_B^{pgT}(a, b + 1)$$

$$P_B^{pgT}(a, b) = p_B P_A^{pgT}(a + 1, b) + q_B P_A^{pgT}(a, b + 1)$$

$$P_A^{pgT}(a, b) = 1, \text{ якщо } a = 7, 0 \leq b \leq 5$$

$$P_B^{pgT}(a, b) = 0, \text{ якщо } b = 7, 0 \leq a \leq 5$$

$$P_A^{pgT}(6, 6) = p_A q_B / (p_A q_B + q_A p_B)$$

$$q_B = 1 - p_B$$

Рекурентні формули для тайбрейку сету матимуть вигляд:

$$P_A^{gst}(c, d) = P_A^{pg}(0, 0) P_B^{gst}(c + 1, d) + [1 - P_A^{pg}(0, 0)] P_B^{gst}(c, d + 1)$$

$$P_A^{gst}(c, d) = 1, \text{ якщо } c = 6, 0 \leq d \leq 4 \text{ або } c = 7, d = 5$$

$$P_A^{gst}(c, d) = 0, \text{ якщо } d = 6, 0 \leq c \leq 4 \text{ або } c = 5, d = 7$$

$$P_A^{gst}(6, 6) = P_A^{pgT}(0, 0)$$

Візуалізація марківської моделі для тай-брейку наведена на рисунку 2.2 [49].

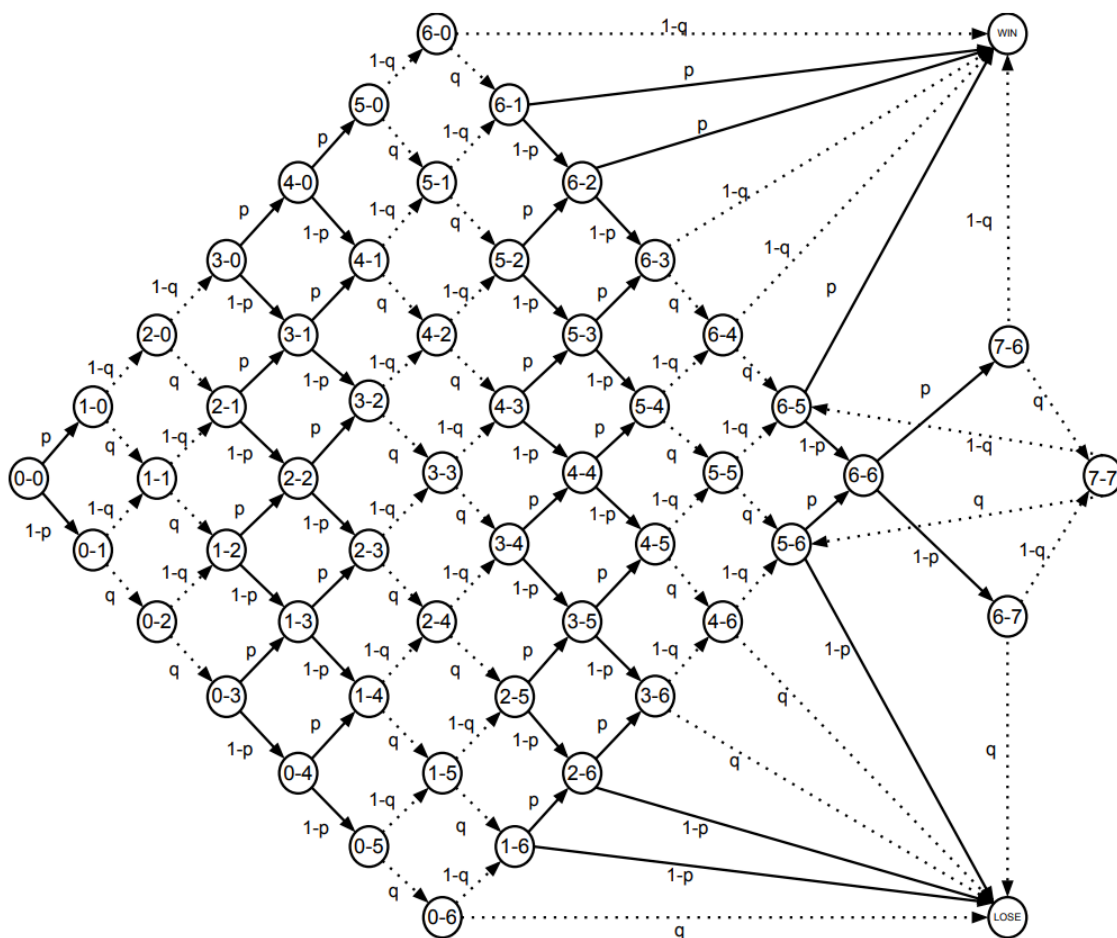


Рисунок 2.2 - Марківська модель тай-брейку в тенісному матчі

## 2.4 Моделювання тенісного сету

Сет або партія — частина матчу в деяких видах спорту, зокрема, у тенісі. Зазвичай матч розбивається на кілька сетів і гра продовжується доти, доки один із супротивників не виграє певну їхню кількість (2-3, залежно від правил гри чи окремого турніру). На відміну від тайму, для якого час гри обмежений, сет грається до певного рахунку. Тенісний сет складається з геймів.

Тенісний сет грається до перемоги одного із супротивників принаймні у 6 геймах. При цьому необхідно, щоб кількість виграних принаймні на два



перевищувала кількість програних геймів. При рахунку 6:6 за геймами більшість сучасних тенісних турнірів передбачає тайбрейк.

Візуалізація марківської моделі для сету наведена на рисунку 2.3 [49].

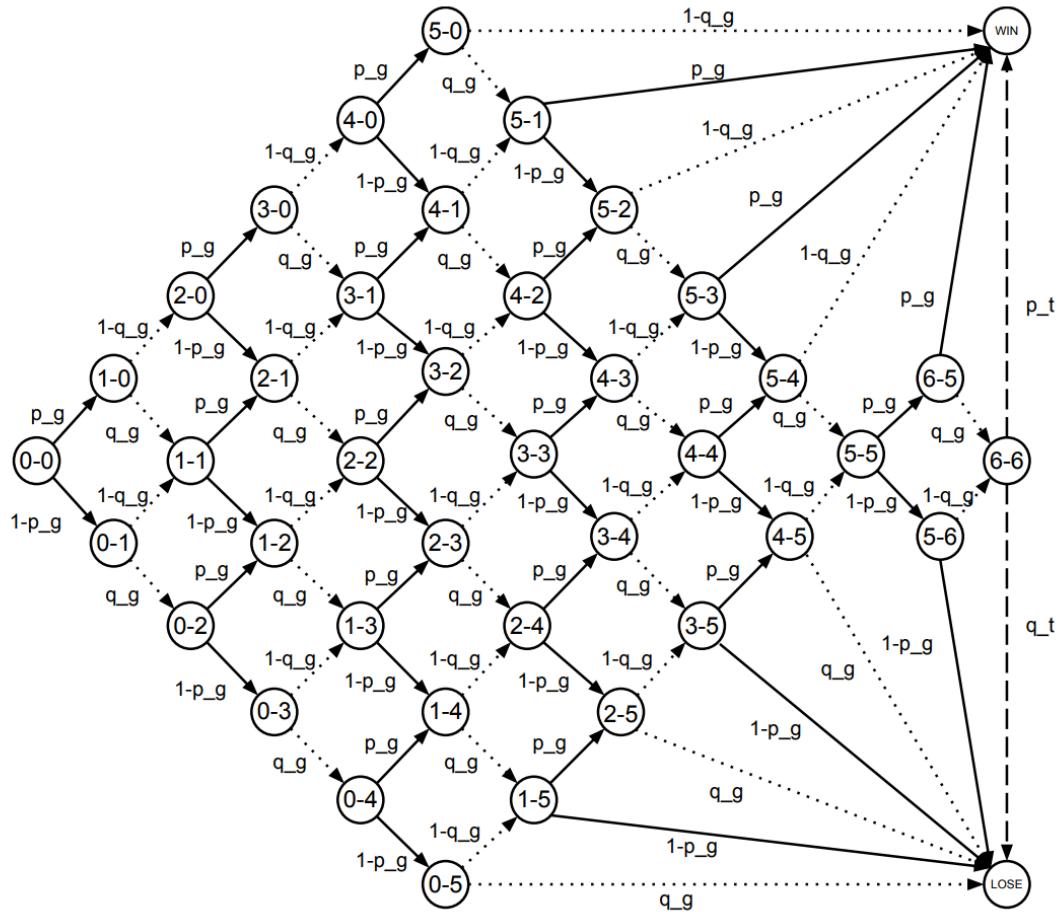


Рисунок 2.3 - Марківська модель сету в тенісному матчі

Нехай  $P_A^{gST}(c, d)$  являється умовною ймовірністю виграшу гравцем  $A$  тай-брейк сету (або за правилом переваги). Тоді при подачі гравцем  $A$  при тай-брейк сеті:

$$P_A^{gST}(c, d) = p_A^g P_B^{gST}(c + 1, d) + (1 - p_A^g) P_B^{gST}(c, d + 1)$$

$$P_A^{gST}(c, d) = 1, \text{ якщо } c = 6, 0 \leq d \leq 4 \text{ або } c = 7, d = 5$$

$$P_A^{gST}(c, d) = 1, \text{ якщо } d = 6, 0 \leq c \leq 4 \text{ або } c = 5, d = 7$$

$$P_A^{gST}(6, 6) = p_A^{gT},$$

де  $p_A^g$  - ймовірність виграшу гравцем  $A$  гейму при подачі,  $p_A^{gT}$  - ймовірність виграшу гравцем  $A$  гейму тай-брейку.

Для сету за правилом переваги запишемо:

$$P_A^{gs}(c, d) = p_A^g P_B^{gs}(c + 1, d) + (1 - p_A^g) P_B^{gs}(c, d + 1)$$

$$P_A^{gs}(c, d) = 1, \text{ якщо } c = 6, 0 \leq d \leq 4$$

$$P_A^{gs}(c, d) = 0, \text{ якщо } d = 6, 0 \leq c \leq 4$$

$$P_A^{gs}(5, 5) = \frac{p_A^g(1 - p_B^g)}{p_A^g(1 - p_B^g) + (1 - p_A^g)p_B^g}$$

## 2.5 Моделювання тенісного матчу

Припустимо, що  $P^{sm}(e, f)$  являє собою умовну ймовірність виграшу гравцем  $A$  матчу за правилом переваги. Тоді запишемо рекурентну формулу:

$$P^{sm}(e, f) = p^{sT} P^{sm}(e + 1, f) + (1 - p^{sT}) P^{sm}(e, f + 1)$$

$$P^{sm}(e, f) = 1, \text{ якщо } e = 3, f \leq 2$$

$$P^{sm}(e, f) = 0, \text{ якщо } f = 3, e \leq 2$$

$$P^{sm}(2, 2) = p^s,$$

де  $p^{sT}$  - ймовірність виграшу гравцем  $A$  тай-брейк сету,  $p^s$  - ймовірність виграшу гравцем  $A$  сету за правилом переваги.

## 2.6 Оцінка ймовірності виграшу при подачі

З огляду на ймовірність того, що обидва гравці виграють очко при їх подачі, ми можемо використовувати ісрархічні вирази, отримані Барнеттом і Кларком, щоб визначити переможця. Залишається питання, як оцінити ці ймовірності виграшу очка при подачі для ще не зіграних матчів. Барнетт і Кларк [63] дають ефективний метод для оцінки цих імовірностей з статистичної статистики гравців:

$$f_i = a_i b_i + (1 - a_i) c_i$$

$$g_i = a_{av} d_i + (1 - a_{av}) e_i,$$

де  $f_i$  - відсоток очок, виграних при подачі гравцем  $i$ ;

$g_i$  - відсоток очок, виграних при прийомі м'яча гравцем  $i$ ;

$a_i$  - відсоток перших подач гравця  $i$ ;

$a_{av}$  - середній відсоток перших подач для всіх гравців;

$b_i$  - відсоток виграшу при першій подачі гравця  $i$ ;

$c_i$  - відсоток виграшу при другій подачі гравця  $i$ ;

$d_i$  - відсоток виграшу при прийомі першої подачі гравцем  $i$ ;

$e_i$  - відсоток виграшу при прийомі другої подачі гравцем  $i$ .

Отже, для матчу між гравцями  $A$  і  $B$  ми можемо оцінити ймовірності виграшу очка при подачі гравцями  $A$  і  $B$  відповідно як  $f_{AB}$  і  $f_{BA}$ , використовуючи наступне рівняння:

$$f_{AB} = f_t + (f_i + f_{av}) - (g_i - g_{av}),$$

де  $f_t$  - відсоток очок, виграних при подачі гравцем  $i$ ;

$f_{av}$  - відсоток очок, виграних при прийомі м'яча гравцем  $i$ ;

$g_{av}$  - відсоток перших подач гравця  $i$ .

Модифікація методу ланцюга Маркова до вирішення задачі прогнозування тенісних матчів може покращити отримувані результати та підвищити точність прогнозу.

## 2.7 Машинне навчання

Машинним навчанням називається розділ штучного інтелекту, що вивчає алгоритми, здатні навчатися або адаптувати свою структуру на підставі обробленої вибірки даних [16]. Машинне навчання з учителем вирішує задачу побудови функції з набору позначених навчальних прикладів, де позначений приклад це пара, що складається з вектора на вході і бажаного значення на виході. Альтернативний спосіб розробки - створення власного програмного засобу, де можна врахувати всі нюанси розв'язуваної задачі, підлаштовувати її під себе. Однак очевидно, що цей шлях вимагає набагато більше зусиль і часу. Але також очевидно і те, що отриманий результат буде точнішим. Природно, щоб піти цим шляхом, потрібно не тільки володіти достатньою кваліфікацією в області штучного інтелекту, а й добре

програмувати самому. Для машинного навчання у прогнозуванні спортивних змагань найчастіше використовують нейронні мережі та логістичну регресію.

Прогнозування за допомогою нейронних мереж має безперечну перевагу над іншими способами прогнозування – це можливість враховувати різноманітні фактори, пов'язані з грою: стан поля і його покриття, виснаження, гравця тощо. Подібні данні можуть скоригувати прогноз та дати точніший результат.

З точки зору існуючих алгоритмів, прогнозування тенісу можна розглянути з двох сторін.

*Регресія* - вихідне значення є важливою величиною. Вихід може представляти ймовірність виграшу матчу безпосередньо, проте справжні ймовірності виграшу минулих матчів невідомі, що змушує нас використовувати дискретні значення для міток навчальних прикладів (наприклад, 1 для перемоги, 0 для ураження). Інакше, можна прогнозувати ймовірність виграшу очка при подачі і вставляти їх в ієрархічні вираження Барнета [63] і О'Майлі [44] для знаходження ймовірності виграшу матчу, але це повертає нас до марківських ланцюгів.

Як завдання бінарної класифікації, в якій можна спробувати *класифікувати матчі за категоріями «перемога» чи «поразки»*. Деякі алгоритми класифікації також дають певний рівень точності події, що належить до класу, що можна використовувати як ймовірність перемоги в матчі.

Запорукою успіху дослідження є правильно підібрані дані – на них нейронна мережа буде вчитися та прогнозувати результати, тобто дані являються навчальними для системи. Навчальні дані потрібно представити таким чином, щоб їх змогла інтерпретувати робоча програма. Важливо врахувати, що не всі наявні вихідні дані повинні в обов'язковому порядку використовуватися в подальшому для кластеризації. Навпаки, змінні результатів матчів принципово повинні бути не навчальними, а використовуватися тільки в ході подальшого кластерного аналізу. Але включати їх до вибірки потрібно, щоб потім по ним можна було виконати кластерний аналіз [16].

Налаштування алгоритму включає пряме або непряме управління розміром мережі, швидкістю навчання і, можливо, в залежності від обраної нейромережевої архітектури, ряд інших параметрів.

Якщо проводиться самонавчання мережі фіксованої структури (мережа Кохонена [65] або на основі радіально-базисних функцій), то мова йде про надання розміру мережі (кількості кластерів). В результаті частина кластерів може виявитися взагалі не задіяними, а деякі переповненими. Незадіяні кластери заважають тільки в процесі навчання мережі, тому що уповільнюють роботу алгоритму. Переповнені кластери набагато небезпечніше, так як їх наявність означає, що нейронна мережа не змогла належним чином розділити навчальні приклади за ознаками, дані будуть усереднені, а точність прогнозу за прикладами, що потрапили в такі кластери, вельми низька.

Якщо ж нейронна мережа не тільки самонавчається, а й самоорганізується, що можливо для всіх раніше перерахованих архітектур, то в підсумку порожніх кластерів просто не буде, а переповненими кластерами в принципі можна управляти, розділяючи їх в ході навчання на більш дрібні по числу вхідних в них прикладів. Зазвичай в якості одного з головних параметрів настройки алгоритму вказується граничне значення відстані від навчального прикладу до центру найближчого кластера. При перевищенні цього порога відповідно до реалізованим алгоритмом створюється новий кластер з центром (ваговими коефіцієнтами), відповідає поточному навчальному наприклад. Однак для регулювання розміру кластера і вирішення проблеми переповненості можна задати ще один параметр - граничне число прикладів, які формують кластер. Якщо виявляється, що новий приклад повинен бути віднесений до того кластеру, ліміт прикладів для якого вже вичерпаний, приймається рішення про створення нового кластера, центр якого буде відповідати наприклад, найбільш віддаленого від центру наявного кластера.

Вибір ще одного налаштування - коефіцієнта швидкості навчання (коефіцієнта адаптації), з одного боку, впливає на швидкість навчання всієї мережі, а з іншого, на міграцію прикладів між прилеглими кластерами. Розумно дотримуватися баланс, щоб він не був занадто великим або занадто малим. Часто для вирішення завдань кластеризації використовується змінний, поступово зменшується коефіцієнт швидкості.

При вдалому налаштуванні і правильній програмній організації алгоритму користувач не бере участі в процедурі навчання. Тривалість навчання залежить як від раніше обраного коефіцієнта швидкості, так і від обсягу навчальних даних.

Кластерний аналіз може бути також виконаний автоматично, якщо програмне забезпечення передбачає відповідну функцію, або проведений вручну. Аналіз складу окремо взятого кластера може включати отримання наступних кількісних характеристик:

- розмір кластера (кількість сформували його прикладів);
- середніх значень навчальних і описують змінних;
- дисперсії цих змінних.

Більшість опублікованих моделей машинного навчання використовують логістичну регресію. Кларк і Дайт [61] застосовують модель логістичної регресії до різниці за очками двох гравців рейтингу ATP для прогнозування результату сету. Іншими словами вони використовують одновимірний простір ознак  $x = (rankdiff)$  і оптимізують  $\beta_1$ , так щоб функція  $\sigma(\beta_1 * rankdiff)$  давала найкращий прогноз для навчальної вибірки. Параметр  $\beta_0$  опущений в моделі, так як  $rankdiff = 0$  дасть можливість перемоги в матчі 0,5. Замість безпосереднього прогнозування результату матчу Кларк і Дає передбачають ймовірність перемоги в сеті і моделюють ймовірність перемоги в матчі, таким чином збільшуючи розмір вибірки. Модель використовувалася для прогнозування результатів декількох чоловічих турнірів в 1998 і 1999 рр., проте автори не наводять даних про точність передбачення, згадуючи лише, що результати задовільні.

Ma, Liu і Tan [60] використовували більший простір ознак з 16 змінних, що належать до трьох категорій: навички і результативність гравця, фізичні параметри гравця, параметри матчу. Модель навчалася на матчах з 1991 по 2008 рр. і використовувалася для видачі рекомендації гравцям: наприклад, «більше тренувати прийом подачі».

Часто в ході машинного прогнозування використовується стохастичне градієнтне спадання. Алгоритм застосовується для ітеративної обробки великих вибірок даних. Нехай є задача пошуку мінімуму  $f(x): R^n \rightarrow R$ , яку можна записати у

вигляді  $f(x) \rightarrow \min_{x \in R^n}$ . Нехай функція  $f(x)$  така, що можна обчислити її градієнт. Тоді можемо використати метод градієнтного спуску для вирішення задачі. Блок-схему методу градієнтного спуску наведено на рисунку 2.4 [2].

### Алгоритм

**Вхід:** функція  $f(x): R^n \rightarrow R$

**Вихід:** знайдена точка оптимуму  $x$

#### 1. Повторювати:

2.  $x^{[k+1]} = x^{[k]} - \lambda^{[k]} \nabla f(x^{[k]})$ , де  $\lambda^{[k]}$  обирається одним зі способів.
3. Якщо виконано критерій зупинки, то повернути поточне значення  $x^{[k+1]}$ .

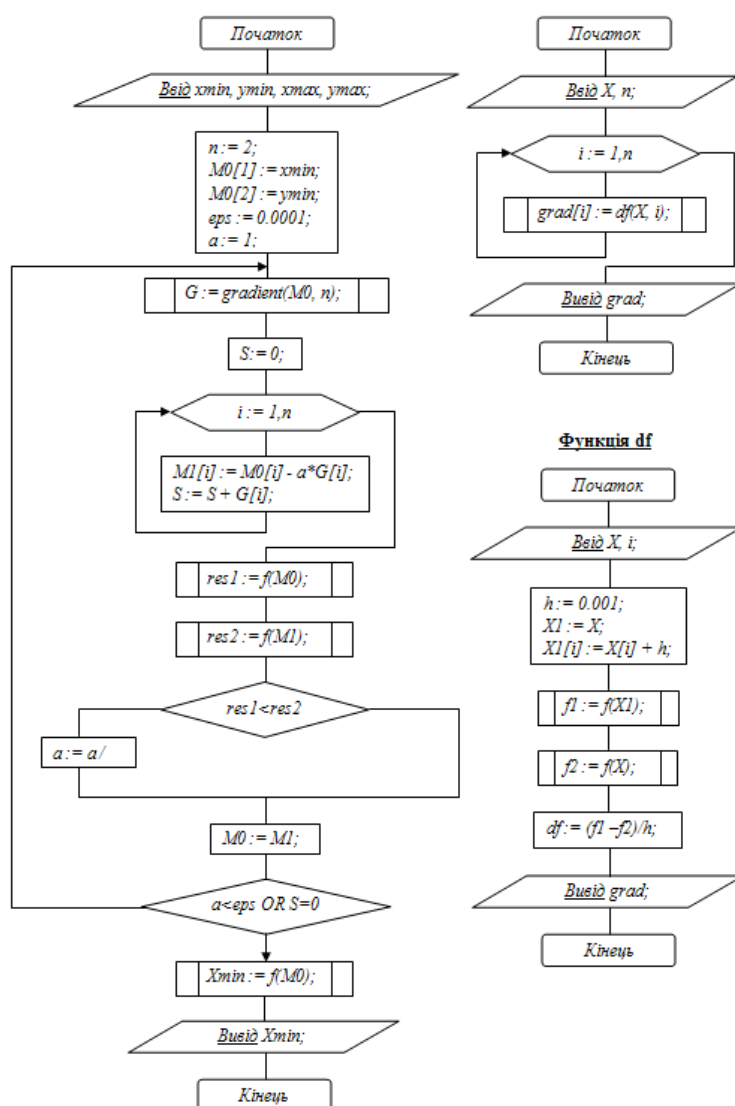


Рисунок 2.4 – Блок-схема методу градієнтного спуску для функції двох змінних

## 2.8 Нечітка логіка

*Нечітка логіка* (англ. *Fuzzy logic*) - розділ математики, що є узагальненням класичної логіки і теорії множин і базується на понятті нечіткої множини, вперше введеного Лютфі Заде в 1965 році як об'єкта з функцією приналежності елемента до множини, що приймає будь-які значення в інтервалі  $[0, 1]$ , а не тільки 0 або 1 [20]. На основі цього поняття вводяться різні логічні операції над нечіткими множинами і формулюється поняття лінгвістичної змінної, в якості значень якої виступають нечіткі множини.

Предметом нечіткої логіки вважається дослідження суджень в умовах нечіткості, розмитості, схожих із судженнями у звичайному сенсі, і їх застосування в обчислювальних системах.

Розглянемо нечітку логіку з точки зору програмування. Тоді нечітка логіка являється набором нестрогих правил, в яких для досягнення поставленої мети можуть використовуватися ідеї, здогадки, а також досвід фахівців, накопичений у відповідній області. Нечіткій логіці властива відсутність суворих стандартів. Найчастіше вона застосовується в експертних системах, нейронних мережах і системах штучного інтелекту. Замість традиційних значень “істинність” і “хибність” в нечіткій логіці використовується більш широкий діапазон значень, серед яких “істинність, хибність, можливо, іноді, ...”. Нечітка логіка відмінно вирішує задачі в тих випадках, коли на поставлене запитання немає чіткої відповіді (так чи ні; «0» або «1») або наперед невідомі всі можливі ситуації. Наприклад, в нечіткій логіці висловлювання виду « $X$  є велике число» інтерпретується як має неточне значення, що характеризується деякою нечіткою множиною. Штучний інтелект і нейронні мережі - це спроба змоделювати на комп'ютері поведінку людини. А так як люди рідко бачать навколишній світ лише в чорно-білому кольорі, виникає необхідність у використанні нечіткої логіки.

Найбільш важливим застосуванням теорії нечітких множин є контролери нечіткої логіки. Їх функціонування дещо відрізняється від роботи звичайних контролерів; для опису системи замість диференціальних рівнянь використовуються



знання експертів. Ці знання можуть бути виражені за допомогою лінгвістичних змінних, які описані нечіткими множинами.

*Лінгвістична змінна* (англ. *Linguistic variable*) — це одне з ключових понять нечіткої логіки [49]. Суть даного поняття полягає в тому, що конкретні значення числової змінної  $x$  зазвичай піддаються суб'єктивній оцінці людини, причому результат такої оцінки виражається природною мовою.

Поняття «лінгвістична змінна» в її спрощеній формі задається у вигляді  $\langle X, T(X), E, G, M \rangle$ , де  $X$  - назва лінгвістичної змінної,  $T(X)$  - множина її значень (терм-множина), що представляють собою найменування нечітких змінних, областю визначення кожної з яких є множина  $X$ ,  $E$  - універсум, тобто весь діапазон значень змінної,  $G$  - синтаксичне правило, що породжує назви нечіткої змінної (тобто терм),  $M$  - семантичне правило, що ставить у відповідність кожній нечіткій змінній  $X$  її значення  $\mu(x)$ .

Загальна структура мікроконтролера, що використовує нечітку логіку. Вона містить у своєму складі наступні складові [20]:

- блок фазифікації;
- базу знань;
- блок рішень;
- блок дефазифікації.

Блок фазифікації перетворює чіткі величини, виміряні на виході об'єкта керування, у нечіткі величини, що описані лінгвістичними змінними в базі знань. Блок рішень використовує нечіткі продукційні правила (ЯКЩО - ТО), закладені в базі знань, для перетворення нечітких вхідних даних у необхідні керуючі впливи, що носять також нечіткий характер. Блок дефазифікації перетворює нечіткі дані з виходу блоку рішень у чітку величину, що використовується для керування об'єктом.

## 2.9 Задача класифікації

Класифікація - один із розділів машинного навчання, присвячений вирішенню наступного завдання. Є множина об'єктів (ситуацій), розділених деяким чином на

класи. Задано скінченну множину об'єктів, для яких відомо, до яких класів вони належать. Це множина називається навчальною вибіркою. Класова приналежність інших об'єктів не відома. Потрібно побудувати алгоритм, здатний класифікувати довільний об'єкт з початкової множини [12].

Класифікувати об'єкт – це означає вказати номер (або найменування класу), до якого належить даний об'єкт. Класифікація об'єкта - номер або найменування класу, що видається алгоритмом класифікації в результаті його застосування до даного конкретного об'єкту.

У машинному навчанні завдання класифікації відноситься до розділу навчання з учителем. Існує також навчання без учителя, коли поділ об'єктів навчальної вибірки на класи не задається, і потрібно класифікувати об'єкти тільки на основі їх подібності між собою. У цьому випадку прийнято говорити про завдання кластеризації або таксономії, і класи називати, відповідно, кластерами або таксонами [1].

Для задачі класифікації можна сформулювати ймовірнісну постановку задачі [0]. Нехай  $X$  - множина описаних об'єктів,  $Y$  - множина номерів (або назв) класів. Існує невідома цільова залежність – відображення  $y^*: X \rightarrow Y$ , значення якої відомі лише на об'єктах кінцевої навчальної вибірки  $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$ , згенерованих відповідно до ймовірнісної міри  $P$ . Необхідно побудувати алгоритм  $a: X \rightarrow Y$ , що може класифікувати довільний об'єкт  $x \in X$ .

Існує типологія задач класифікації. За типом вхідних даних ці задачі поділяють на [10]:

- *характеристичний опис* - найбільш поширений випадок. Кожен об'єкт описується набором своїх характеристик, які називаються ознаками. Ознаки можуть бути числовими або нечисловими.
- *матриця відстаней між об'єктами*. Кожен об'єкт описується відстанями до всіх інших об'єктів навчальної вибірки. З цим типом вхідних даних працюють деякі методи, зокрема, метод найближчих сусідів, метод парzenовського вікна, метод потенційних функцій.

- *часовий ряд або сигнал* являє собою послідовність вимірювань в часі. Кожен вимір може представлятися числом, вектором, а в загальному випадку - характеристична описом досліджуваного об'єкта в даний момент часу.
- *зображення або відеоряд*.
- зустрічаються і більш складні випадки, коли вхідні дані подаються у вигляді *графів, текстів, результатів запитів до бази даних тощо*. Як правило, вони наводяться до першого або другого випадку шляхом попередньої обробки даних і вилучення ознак.
- класифікацію сигналів і зображень називають також *розпізнаванням образів*.

За типами класів задачу класифікації поділяють на наступні види [11]:

- *двокласова класифікація*. Найбільш простий в технічному відношенні випадок, який служить основою для вирішення більш складних завдань.
- *багатокласова класифікація*. Коли число класів досягає багатьох тисяч (наприклад, при розпізнаванні ієрогліфів або злитого мовлення), завдання класифікації стає істотно більш важкою.
- *класи, що не перетинаються*.
- *класи, що перетинаються*. Об'єкт може належати одночасно до кількох класів.
- *нечіткі класи*. Потрібно визначати ступінь належності об'єкта кожному з класів, зазвичай це дійсне число від 0 до 1.

## 2.10 Висновок до розділу

У даному розділі було розглянуто існуючі підходи до прогнозування результатів спортивних змагань та тенісних матчів. Основними підходами до прогнозування являються стохастичні та машинні.

Стохастичним розглянутим методом являється моделювання тенісного матчу за допомогою ланцюгів Маркова. Такий підхід будує загальну модель матчу та на основі теорії про незалежний однаковий розподіл показує умовірність виграшу в кожен момент матчу. Однак, такий спосіб прогнозування не може давати більш розширену

інформацію про ймовірність виграшу, враховуючи фізичні параметри гравця, покриття корту тощо.

Машинним навчанням називається розділ штучного інтелекту, що вивчає алгоритми, здатні навчатися або адаптувати свою структуру на підставі обробленої вибірки даних. Прогнозування за допомогою нейронних мереж має безперечну перевагу над іншими способами прогнозування – це можливість враховувати різноманітні фактори, пов'язані з грою: стан поля і його покриття, виснаження, гравця тощо. Подібні данні можуть скоригувати прогноз та дати точніший результат.

У машинному навчанні завдання класифікації відноситься до розділу навчання з учителем. Існує також навчання без учителя, коли поділ об'єктів навчальної вибірки на класи не задається, і потрібно класифікувати об'єкти тільки на основі їх подібності між собою. У цьому випадку прийнято говорити про завдання кластеризації або таксономії, і класи називати, відповідно, кластерами або таксонами.

### **3 РОЗРОБКА МЕТОДУ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ СПОРТИВНИХ ЗМАГАНЬ**

#### **3.1 Ідея запропонованого алгоритму**

На сьогоднішній день теніс являється одним із найпопулярніших видів спорту на планеті, що й зумовлює постійні модифікації та розробки сучасніших та точніших алгоритмів прогнозування результатів. Алгоритми базуються на різноманітних підходах, однак основні методи сьогодення – ланцюги Маркова та нейронні мережі.

Прогнозування спортивних змагань пройшло довгий шлях модифікацій, перш ніж змогло постати перед нами в тому вигляді, якому ми можемо бачити його зараз. Дане дослідження використовує як традиційні стохастичні методи прогнозування результатів тенісних турнірів – ланцюги Маркова, так і новітні технології – нейронні мережі. Комбінування стохастичних методів та машинних можуть давати методологію, за допомогою якої можливе утворення нового підходу до прогнозування результатів та визначення факторів, що впливають на результативність гравця. Тож розглянемо обидві категорії методів прогнозування та, промодифікувавши один зі стохастичних підходів і зкомбінувавши їх у єдину методологію, запропонуємо відповідні модифікації для покращення фінальних результатів.

Одним з кроків методології дослідження являється прогнозування на існуючих статистичних даних змагань за допомогою ієрархічних стохастичних моделей, заснованих на ланцюгах Маркова. Стохастичні моделі призначені для аналізу та прогнозування явищ в умовах невизначеності вихідних даних і реалізуються методами математичної статистики. Тобто, це модель, де враховуються випадкові фактори. Такий апарат є базовим для сучасних моделей прогнозування тенісу.

Прогнозування результатів тенісного турніру на основі статистичних даних у даному дослідженні проводиться також за допомогою машинного навчання, тобто нейронних мереж. Нейромережевий підхід не містить в собі модельних обмежень, він підходить як для лінійних і складних нелінійних задач, а також завдань класифікації. Навчання нейронної мережі в першу чергу полягає в зміні «сили» зв'язків між

нейронами. На початку побудови та дослідження необхідно вказувати змінні, які аналізуються і передбачаються з достатнім рівнем деталізації. На використовуваний рівень деталізації впливає безліч факторів: доступність і точність даних, вартість аналізу і переваги користувачів результатів прогнозування. У ситуаціях, коли найкращий набір змінних неясний, можна спробувати різні альтернативи і вибрати один з варіантів, який дає найкращі результати.

Даний розділ показує етапи новоствореного алгоритму прогнозування результатів тенісного матчу. На основі наведеного алгоритму стало можливим комбінування ієрархічного стохастичного та машинного методів прогнозування результатів спортивних змагань. Він включає в себе наступні загальні етапи:

- збір даних про матчі, проведені гравцем раніше;
- заповнення ймовірнісної таблиці даними гравця;
- побудову ланцюгів Маркова на основі ситуацій, коли гравець або подає першим, або приймає першим (побудова модифікованим методом), або є в ситуації тай-брейку;
- аналіз даних за допомогою машинного навчання для виявлення головних факторів впливу на результат матчу (виграш чи програш гравця);
- оновлення даних на основі знайдених головних факторів;
- повторну побудову моделі матчу за допомогою ланцюгів Маркова для проведення фінального прогнозу;
- порівняння точності прогнозу із існуючими методами та алгоритмами.

Блок-схему алгоритму наведено на рисунку 3.1.

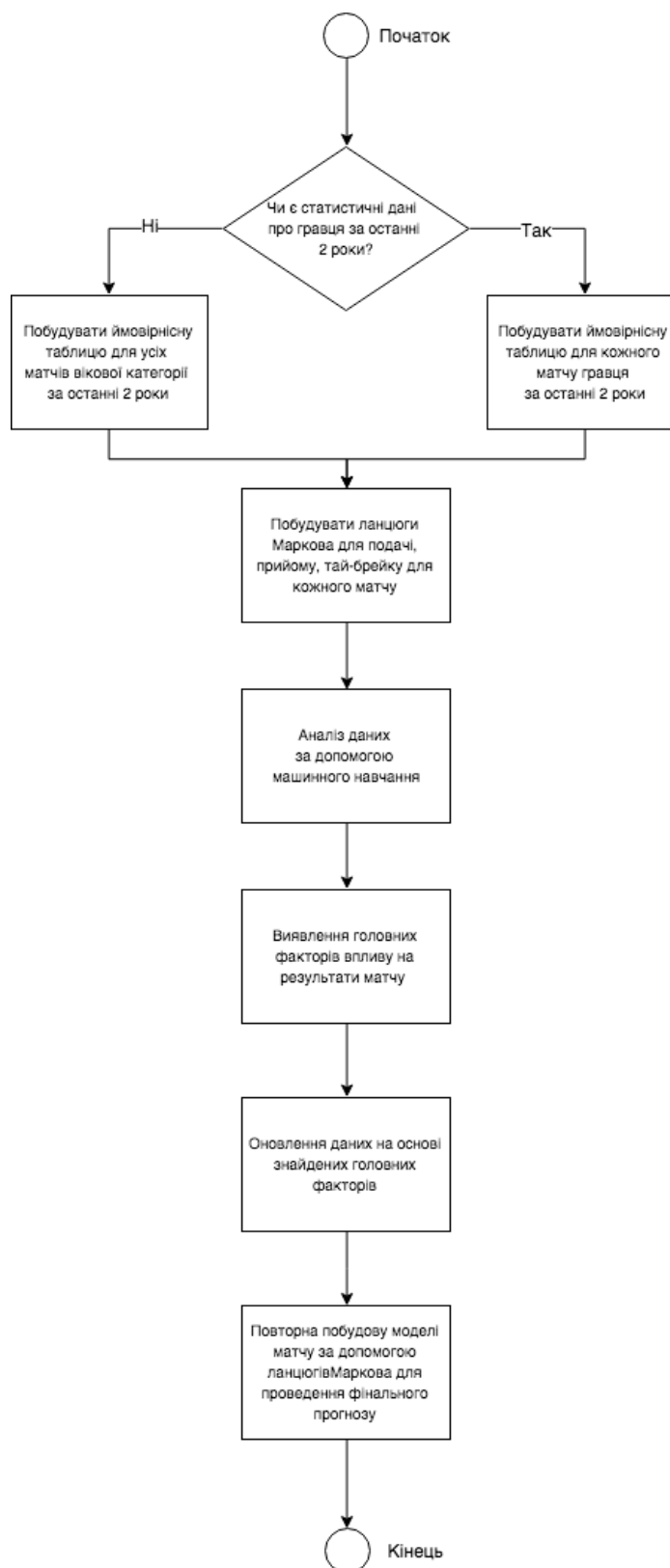


Рисунок 3.1- Блок-схема алгоритму прогнозування результатів тенісного матчу

### 3.2 Формальна постановка задачі

Припустимо, що маємо гравця  $k$ , для якого буде побудовано прогнозування. Якщо гравець має попередню статистику, то, відповідно, беремо статистичні дані за період довжиною в 2 роки –  $m$  ігор. Якщо гравець не володіє подібною статистикою, то аналіз будується на основі моделі “узагальненого” гравця, яка вибудовується з усіх наявних даних для гравців такого ж віку, як гравець  $k$  за період тривалістю 2 рок –  $m$  ігор. Існує деякий вектор факторів  $\{X\} = \{x_1, \dots, x_n\}$ , що визначає умови, за яких відбувалася гра.

Для деякої  $j$ -ої гри,  $j = \overline{1, m}$  маємо:

- $p_j^1$  - ймовірність здобуття очка при умові, що гравець  $k$  подає в геймі першим;
- $p_j^2$  - ймовірність здобуття очка при умові, що гравець  $k$  приймає в геймі першим;
- $p_j^T$  - ймовірність здобуття очка гравцем  $k$  в тай-брейку.

Існує деякий вектор  $y_j$  результатів матчу  $j$ -ої гри,  $j = \overline{1, m}$ , такий, що:

$$y_j = \begin{cases} 1, & \text{якщо гру було виграно;} \\ 0, & \text{якщо гру було програно.} \end{cases}$$

Для зручності обчислень та наглядності результатів прогнозування введемо ймовірнісну таблицю, де розміщуватимемо статистичні дані про матчі та дані про матч, що буде прогнозуватися.

На основі таких вхідних даних є можливою побудова ймовірнісної таблиці, що наведена в таблиці 3.1.



Таблиця 3.1 – Ймовірнісна таблиця для вхідних даних прогнозування

Гра	$x_1$	...	$x_n$	$p_j^1$	$p_j^2$	$p_j^T$	$y_j$
1	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
$m$	...	...	...	...	...	...	...

### 3.3 Модифікований метод ланцюгів Маркова у прогнозуванні тенісного матчу

Ланцюг Маркова - це випадковий процес, що задовольняє властивість Маркова і який приймає скінченну чи зліченну кількість значень (станів). Випадкова послідовність подій називається ланцюгом Маркова, якщо кожен перехід з одного стану в інший не залежить від того, коли і як система прийшла у поточний стан. Початковий стан може бути заданим заздалегідь або бути випадковим. Необхідно звернути увагу на те, що наступний стан залежить лише від поточного стану системи, усі попередні стани не враховуються - основуючись на цьому твердженні, можна побудувати ланцюг Маркова для тенісного матчу.

Імовірність виграшу очка у професійному тенісі являється ключовою у більшості способів прогнозування. Дана величина забезпечує обчислювальну базу, на якій засновується розрахунок ймовірності виграшу матчу.

### 3.3.1 Застосування ланцюгів Маркова в ході алгоритму

У ході алгоритму застосовується побудова ланцюгів Маркова як за допомогою модифікованого методу, так і для методу побудови для ситуації гейму при подачі гравцем, а також при ситуації тай-брейку.

Згідно з формальною постановкою задачі (пункт 3.2) нижче буде наведено структуру побудови ланцюгів для тенісного матчу. Для деякої  $j$ -ої гри,  $j = \overline{1, m}$  з урахуванням структури тенісної матчу (2 сету, кожен з яких складається з геймів та при рахунку 6-6 у сеті – ситуації тай-брейку) можна побудувати ймовірнісні моделі для сету на основі ланцюгів Маркова, зі структурою, наведеною на рисунку 3.2.

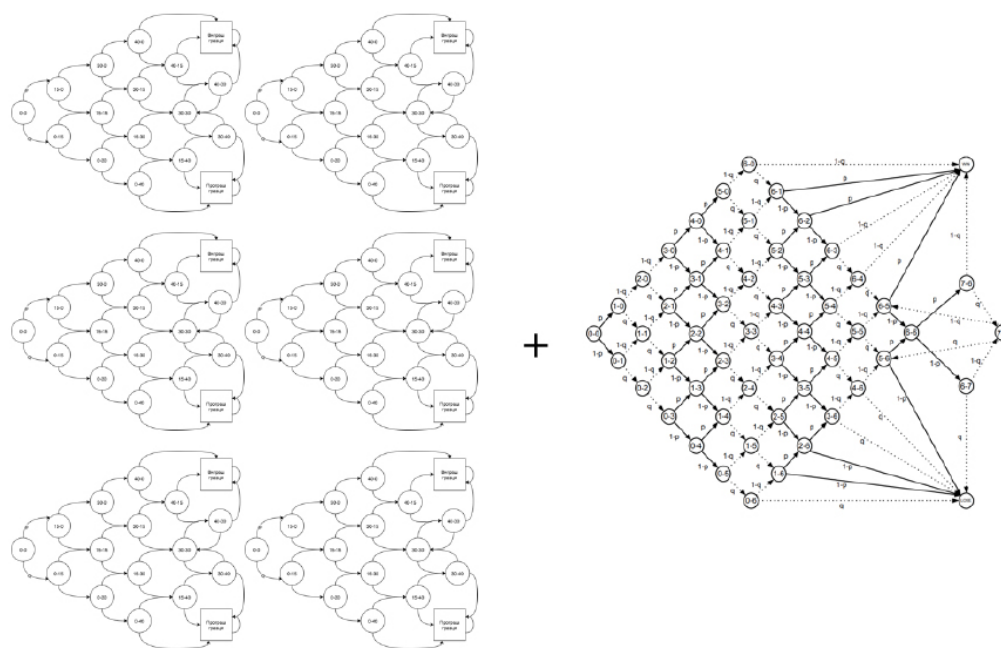


Рисунок 3.2 – Модель сету для тенісного матчу в загальному випадку

### 3.3.2 Математична постановка

Модифікований метод ланцюгів Маркова будується за умови, що для обчислення використовується  $p_j^2$  - ймовірність здобуття очка при умові, що гравець  $k$

приймає в геймі першим для деякої  $j$ -ої гри,  $j = \overline{1, m}$ . Таким чином, ми отримуємо інверсований метод обчислення ймовірності виграшу очка в геймі.

Припустимо, що ми розглядаємо матч, у якому приймають участь два гравці тенісного матчу, перший гравець -  $A$ , другий гравець -  $B$ . Припустимо, що  $p_j^2$  - ймовірність того, що гравець  $A$  виграє очко, якщо подаватиме  $B$ . Відповідно, припустимо, що  $q = 1 - p_j^2$  - ймовірність того, що гравець  $B$  виграє очко, якщо подаватиме. Нехай ймовірність виграшу очка незалежно та однаково розподілена.

Така постановка дає нам можливість побудувати ланцюг Маркова, за допомогою якого буде описано ймовірність виграшу гравця в геймі.

Очки матчу є незалежними й однаково розподіленими, тобто ймовірність виграшу очка константна протягом матчу.

### 3.3.3 Допустимі стани та множина пошуку рішень

Побудуємо марківський процес із дискретним часом, кінцевим станом та матрицею переходів  $M$  і початковим станом  $I_0$ . Множина станів цього ланцюга являється поглинаючою (або замкненою), так як включає в себе 2 стани, з яких не може бути переходу в наступний стан – це стан виграшу гравця  $A$  та стан виграшу гравця  $B$  [71]. Оскільки ми модифікуємо ситуацію гри гейму, то матриця  $M$  набуває вигляду:

$$M = \begin{bmatrix} m_{11} & \cdots & m_{117} \\ \vdots & \ddots & \vdots \\ m_{171} & \cdots & m_{1717} \end{bmatrix}$$

Це зумовлено тим, що при моделюванні марківського ланцюга для гейму гри може бути всього 17 станів системи. Запишемо загальний вигляд блочної структури даної матриці:

$$M = \begin{bmatrix} E & 0 \\ R & W \end{bmatrix}$$

Тут  $E$  - одинична підматриця, порядок якої збігається з числом поглинаючих станів;  $W$  - квадратна підматриця ймовірностей переходів на множині неповоротних станів;  $R$  - прямокутна підматриця переходів з неповоротних станів в поглинаючі;  $0$  -

нульова підматриця. Якщо число загальних станів – 17, з них 2 – поглинаючі стани, тоді підматриці мають наступний порядок:

$$M_{17 \times 17}, E_{15 \times 15}, W_{2 \times 2}, R_{2 \times 15}, O_{15 \times 2}$$

Візуалізацію ланцюга Маркова за умови, що приймає подачу гравець  $k$ , наведено на рисунку 3.3:

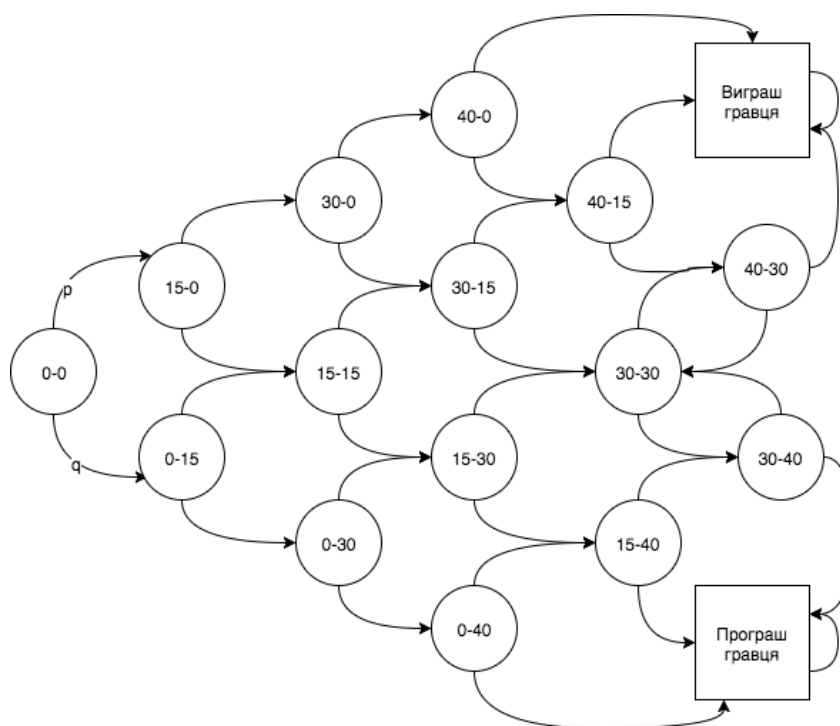


Рисунок 3.3 - Марківська модель гейму в матчі за умови, що подає гравець  $k$

Використання марківських ланцюгів для прогнозування результатів тенісних турнірів не можна вважати досконалим методом, оскільки при цьому абсолютно ігнорується безліч факторів, що впливають на гру: самопочуття гравця, наявність нещодавніх травм, графік тренування, тип покриття корту тощо. Однак, при комбінуванні методів та створенні нових алгоритмів стає можливим точніше визначати ймовірність виграшу гравця [73].

### 3.3.4 Алгоритм проходу по ланцюгу гейму

При створенні модифікованого алгоритму необхідно означити його кроки, вхідні та вихідні дані, умови кінця алгоритму. Нижче наведено алгоритм проходу по ланцюгу гейму для гравця  $k$  [72].

### **Алгоритм проходження по ланцюгу гейму**

**Вхід:** ймовірність виграшу очка гравцем  $k$  при прийомі  $p_j^2$

**Вихід:** ймовірність виграшу гейму гравцем  $k$

**1. Проініціалізувати стани системи.**

**2. Для станів системи повторювати:**

2.1  $P_k(a, b) = p_j^2 P_k(a + 1, b) + (1 - p_j^2) P_k(a, b + 1)$ , де  $a, b$  – очки, отримані кожним з гравців за обмежень  $P_k(a, b) = 1$ , якщо  $a = 4, b \leq 2$ ;  $P_k(a, b) = 0$ , якщо  $b = 4, a \leq 2$

**3. Якщо пройдено усі стани системи, то завершити.**

Так ітеративно проходить обчислення для кожного стану моделі.

### **3.4 Прогнозування тенісного матчу за допомогою нейронних мереж**

Машинним навчанням називається розділ штучного інтелекту, який вивчає алгоритми, здатні навчатися або адаптувати свою структуру на підставі обробленої вибірки даних. Машинне навчання може бути з учителем або без учителя. Навчання з учителем - один із способів машинного навчання, в ході якого випробувана система примусово навчається за допомогою наявної множини прикладів «стимул-реакція» з метою визначення «реакції» для «стимулів», які не належать наявній множині прикладів.

В нашому випадку навчальною вибіркою може служити наявна статистика результатів тенісних матчів.

У пункті розглянуто теоретичні аспекти побудови прогнозу та нейромережевого підходу, в основі якого лежать запропоновані методи прогнозу, а також деталі фундаментальні засади їх побудови і реалізація [70].

#### **3.4.1 Вплив покриття корту на результат прогнозування**

Історично склалося так, що покриття перших тенісних змагань було трав'яним. Сьогодні найчастіше використовується три види покриття: жорсткі (із використанням

синтетики), глиняні, і трав'яні. Кожна з цих поверхонь – по різному впливає на швидкість гри, віддачу м'яча і рухи гравців. Це стає умовою для особливості фізичних потреб, техніки та тактики, що потребує гнучкості гравців і відображається на продуктивності. Саме тому необхідно враховувати фактор покриття при прогнозуванні.

Це значний фактор для побудови прогнозування, так як різні поверхні вимагають різних функцій від спортсменів. Наприклад, на траві гравці подають краще і здатні на агресивну гру і, як правило, мають на цій поверхні найкращу продуктивність. Навпаки, на глині, повільнішій поверхні, як правило, найкраще проявляються оборонні прийоми гравця.

Аналіз в базі даних, використаний у цій роботі, приводить до висновку, що для прогнозування необхідно вивести кореляцію між поверхнями покриття. Даний розрахунок показує міцність лінійного співвідношення між двома типами покриття, і він розраховується за допомогою рівняння, що являє собою коефіцієнт Пірса (2) [48]:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[ \sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[ \sum_{i=1}^n (y_i - \bar{y})^2 \right]}}$$

де  $x$  та  $y$  – вхідні вектори, що мають  $n$  значень кожен. Значення  $r$ , що наближаються до 1, вказують на сильні лінійні зв'язки, а нульові значення вказують на відсутність лінійного зв'язку між векторами.

### 3.4.2 Застосування нечіткої логіки в процесі прогнозування

Визначною роботою у розгляді даного питання являється стаття Заде [69] - нечітка логіка використовується для вирішення різноманітних поставлених задач і являється Системою Висновувань (Inference Systems). Введені Мамдані, такі системи керуються апроксимованими міркуваннями, відомими як узагальнений Modus Ponens. Він заснований на лінгвістичних змінних та ЯКЩО-ТО імплікативних правилах для формування типових висновувань нечітких систем, при цьому

використовуючи людський досвід для розвитку інтелектуальних алгоритмів, здатний справлятися з неоднорідними/неточними даними.

Побудова нейронної мережі заснована на статті Фернандеса [48]. У нашому випадку застосування базується на нульовому порядку Сугено системи виведення [62], де правила імплікації є постійними. Нечіткий предиктор використовує дані про матчі на вході. Для кожного з матчів ці входні дані обчислюються, проходять процес фазифікації та розбиваються на категорії впливу, задані імплікативними правилами типу ЯКЩО-ТО за допомогою функції, визначеної для загального випадку за допомогою (3) [48]:

$$df(x; a; b; c) \begin{cases} 0, x \leq a \text{ або } c \leq x \\ \frac{x - a}{b - a}, a \leq x \leq b \\ \frac{c - x}{c - b}, c \leq x \leq b \\ 0 \end{cases}$$

Характеристики для побудови даної функції наведено на рисунку 3.4 [48].

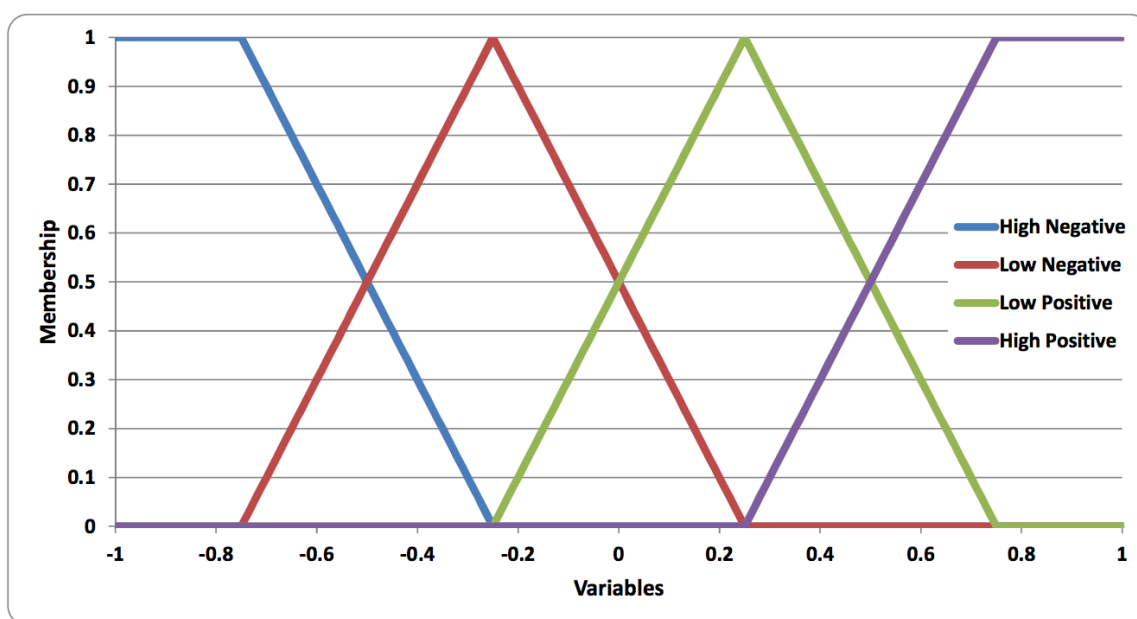


Рисунок 3.4 - Функції членства, що використовуються нечітким предиктором

Трикутні функції членства вибираються через їх математичну простоту та ефективність, в результаті чого зменшуються обчислювальні витрати. При використанні більш складного апарату (наприклад, функцій Гауса) не спостерігалось

покращень значень прогнозу. Також не було відмічено покращень при оптимізації кількості класів або їх параметрів для точної настройки.

Для генерації набору правил, що визначають Систему Висновувань (Inference System), основним джерелом інформації є досвід людини. Тут fuzzy-логіка показує її основне застосування - вираження математичних знань, які зазвичай розглядаються в лінгвістичній формі. Таким чином, встановлення двох можливих результатів - перемога або програш гравця – будують базу правил для аналізу змінних. Зразок деяких з цих правил показано у таблиці 3.2 [48]. Оператори AND (I) виконуються за функцією відповідно до [48]:

$$\mu_C(x) = \min(\mu_A(x), \mu_B(x)),$$

де  $\mu_A(x), \mu_B(x)$  - обрані функції членства.

Таблиця 3.2 – Приклад бази правил для системи

Правила	Ранг		Статистика		Покриття		Результат
ЯКЩО	високий позитивний	I	високий позитивний	I	-	ТО	виграш
ЯКЩО	низький позитивний	I	низький позитивний	I	низький позитивний	ТО	виграш
ЯКЩО	високий негативний	I	високий негативний	I	високий негативний	ТО	виграш
ЯКЩО	високий позитивний	I	високий позитивний	I	високий позитивний	ТО	програш
ЯКЩО	низький негативний	I	низький негативний	I	низький негативний	ТО	програш
ЯКЩО	високий негативний	I	високий негативний	I	-	ТО	програш



Описаний метод дає вагові коефіцієнти для статистичних даних. Їх значення розподіляються в діапазоні  $[0;1]$  та за допомогою віднесення до класу значущості можуть бути визначені як впливові.

Система виводу задачі наведена на рисунку 3.5.



Рисунок 3.5 – Система виводу задачі

### 3.4.3 Застосування нейронних мереж в процесі прогнозування

Є безліч методів для формування архітектури нейронної мережі. Найпоширенішим являється багатошаровий перцептрон (Multi Layer Perceptron, або MLP). Це мережа з прямим поширенням сигналу, де нейрони розташовуються в послідовних шарах, а виходи кожного нейрона в кожному шарі підключаються до входів наступного шару. Інформація мережі зберігається у вагах, пов'язаних з кожним з цих з'єднань, причому їх "навчання" здійснюють ітераційними алгоритмами, які коригують ці ваги на основі прикладів (пара входів-виходів, відома апріорно).

Багатошаровий перцептрон з одним проміжним шаром і одним вихідним шаром здатний вирішувати деякі нелінійні задачі та наближати неперервні функції, а додавання ще одного проміжного шару дозволяє йому реалізувати будь-яку функцію, незалежно від того, чи має вона лінійний розподіл, як показав Цібенко [40]. Кількість вузлів у кожному шарі – основний фактор, що відповідає за конвергенцію (формування схожих ознак у неспоріднених груп) на етапі навчання та для точності результатів [36].

Кожен із вузлів нейронної мережі містить функцію активації, яка відповідає за обчислення виходу вузла з вагової суми його входів. Найбільш розповсюдженими функціями активації є ті, що базуються на сигмоїдальних функціях, бо за допомогою них можна визначати баланс між лінійною та нелінійною поведінкою, а також тому, що вони є безперервними монотонно зростаючими функціями, диференційованими в усіх точках [42].

### 3.4.4 Застосування логістичної регресії та гіперболічної функції дотичної активації в процесі прогнозування

Обрана для прогнозування сигмоїдальна функція являється логістичною регресією. Логістична регресія або логіт-регресія — статистичний регресійний метод, що застосовують у випадку, коли залежна змінна є категорійною, тобто може набувати тільки двох значень (чи, загальніше, скінченної множини значень).

Логістична регресія - це по суті алгоритм класифікації [17]. Головними в алгоритмі є властивості логістичної функції. Логістична функція  $\sigma(t)$  визначається як:

$$\sigma(t) = 1/(1 + e^{-t})$$

Як видно на рисунку 3.6 [17], логістична функція відображає суттєві входні значення в діапазоні від  $-\infty$  до  $+\infty$  і від 0 до 1, дозволяючи інтерпретувати виходи як ймовірності.

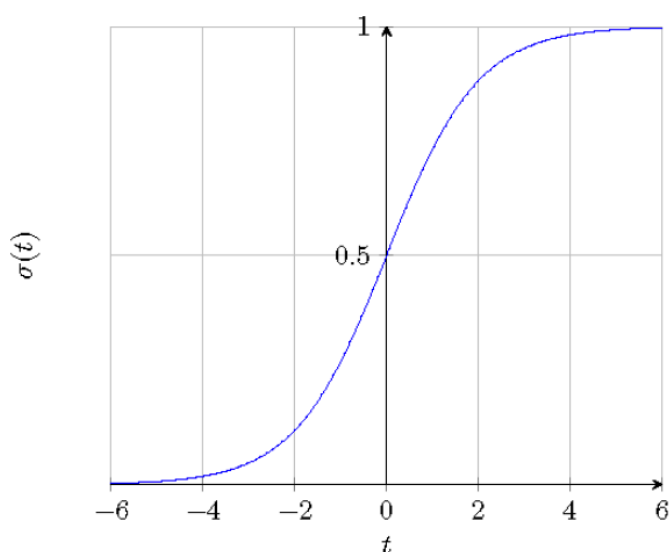


Рисунок 3.6— Логістична функція

Модель логістичної регресії [17] для прогнозування матчів складається з вектора  $n$  ознак матчу  $x = x_0, x_1, \dots, x_n$  і вектора  $n + 1$  параметрів моделі  $\beta = \beta_0, \beta_1, \dots, \beta_n$ . Для прогнозування за допомогою моделі спочатку проектуємо точку в нашому  $n$ -розмірному просторі ознак на дійсне число:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Тепер можна перетворити  $z$  в значення в прийнятному діапазоні ймовірностей (від 0 до 1) за допомогою логістичної функції, визначеної вище:

$$p = \sigma(z) = 1/(1 + e^{-z})$$

Навчання моделі складається з оптимізації параметрів  $\beta$ , так щоб модель давала найкраще відтворення результатів матчів для навчальної вибірки. Це здійснюється шляхом мінімізації функції логістичних втрат (рівняння нижче), яка дає міру похибки моделі при прогнозуванні результатів матчів, що використовувалися для навчання.

Рисунок 3.7 [17] нижче показує логістичні втрати, що виникають через одного матчу для різних прогнозованих ймовірностей, за умови, що матч завершився перемогою передбаченого гравця. Будь-яке відхилення від найточнішого передбачення  $p = 1,0$  штрафнується.

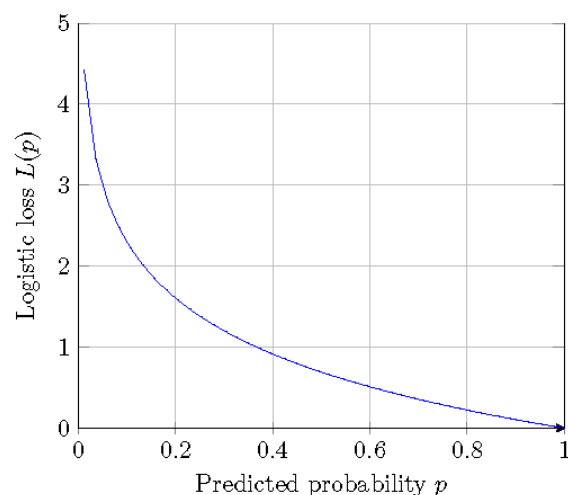


Рисунок 3.7 – Логістичні втрати при правильному прогнозі матчу

Гіперболічна функція дотичної активації використовується безпосередньо в ході прогнозування та нейромережевого підходу [17].

$$f(x) = \tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$$

Наведені функції представляють властивість стискання вхідних даних, причому асимптотично великі додатні значення до значення 1 і великі від'ємні значення прирівнюються до 0.

### 3.4.5 Ідея роботи методу із застосуванням нейронної мережі

У цьому дослідженні нейронна мережа навчається на наборі даних про матчі, де переможці вже відомі як дані з історії та попередніх виступів гравців, із можливістю прогнозувати переможця у нових матчах при наданні оновлених даних [48].

Дана нейронна мережа складається з вхідного шару з чотирма нейронами і гіперболічної функції дотичної активації, прихованого шар з чотирма нейронами і функції логістичної активації, а також вихідного шару з нейрону і гіперболічної дотичної активаційної функції. Всі зв'язки між вузлами зважуються на питому вагу:  $W_i$  - вектор для вхідного шару,  $W_h$  - для проміжного (прихованого) шару і  $W_o$  - для виводу. Як вхідні змінні для нейронної мережі, подається три групи факторів, на яких ґрунтується нечіткий предиктор: рейтинг, статистичні та фізичні параметри гравця та поверхня корту. На виході ми отримуємо значення ваг факторів, що впливають на результативність гравця. Всі змінні нормуються для інтервалу [0, 1].

Архітектуру нейронної мережі наведено на відповідному рисунку 3.8 [48].

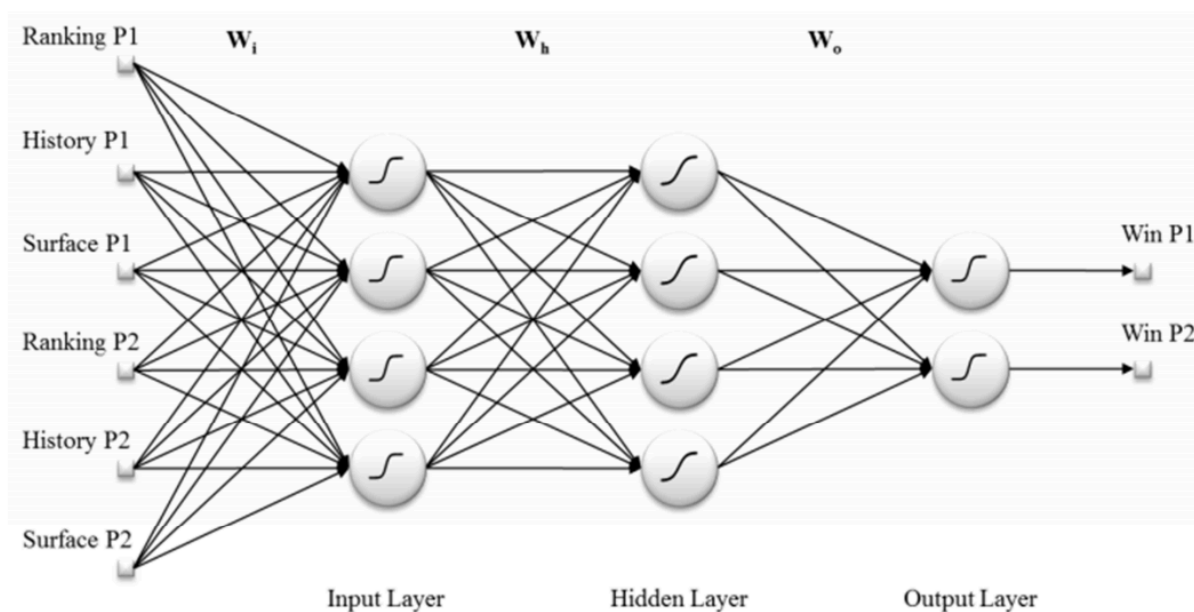


Рисунок 3.8 – Архітектура нейронної мережі

Навчання нейронної мережі засноване на алгоритмі навчання, відомому як зворотне поширення. У цьому методі маси зв'язків між вузлами мережі ініціалізуються випадковими значеннями. Після цього набори вхідних значень, які вже приводять до виводу, вже представлені в мережі (у випадковому порядку). Для кожної з цих множин, вихід мережі з поточними вагами обчислюється, в так званій "прямій фазі". Отриманий результат потім порівнюється з правильним вихідним шаблоном, щоб розрахувати помилку між ними. Ця помилка поширюється через мережу в зворотний шлях ("зворотна фаза", що виправдовує назву алгоритму).

Результуючи похибки кожного виходу, постійна "швидкість навчання" віднімається від зв'язків "ваги" відповідного вузла в останньому шарі. Помилка в кожному вузлі попередніх шарів розраховується з використанням помилок вузлів з наступного, пов'язаного з ним шару, зважена вагами зв'язків між ними.

### 3.5 Алгоритм фінального прогнозування для гравця

У даному пункті наведено апарат, за допомогою якого кобінуються методи, описані у пунктах 3.3 та 3.4 та розраховуються фінальні результати прогнозування тенісного матчу.

### 3.5.1 Алгоритм вибору ваг факторів

Для розрахунку фінальних ймовірностей необхідно знати, які ж саме фактори являються значущими для прогнозування тенісного матчу. У пункті 3.4 було описано, як знаходяться вагові коефіцієнти факторів. На поточному кроці прогнозування нам необхідно обрати ті, що задовольняють нашому матчу і розрахувати коефіцієнт для уточнення ймовірності вдалої подачі, прийому чи тай-брейку.

Тоді кожного фактора  $x_i$ ,  $i = \overline{1, n}$ , що відповідає вектору факторів  $\{X\} = \{x_1, \dots, x_n\}$  з таблиці (Таблиця 3.1) необхідно визначити належність загальних факторів  $\{X^h\} = \{x_1^h, \dots, x_n^h\}$  до цієї вибірки та обрати відповідні ваги факторів  $\{W^x\} = \{w_1^x, \dots, w_n^x\}$ , сформувавши вектор ваг факторів  $\{W^{x'}\} = \{w_1^{x'}, \dots, w_n^{x'}\}$  для поточного матчу та ваговий коефіцієнт  $k$  ( $0 \leq k \leq 1$ ).

#### Алгоритм вибору ваг факторів

**Вхід:** вектор загальних факторів  $\{X^h\} = \{x_1^h, \dots, x_n^h\}$ , вектор факторів поточного матчу  $\{X\} = \{x_1, \dots, x_n\}$ , вектор ваг факторів  $\{W^x\} = \{w_1^x, \dots, w_n^x\}$

**Вихід:** вектор ваг факторів поточного матчу  $\{W^{x'}\} = \{w_1^{x'}, \dots, w_n^{x'}\}$ , ваговий коефіцієнт  $k$

**1. Для кожного  $i$ ,  $i = \overline{1, n}$ :**

$$1.1 \quad w_i^{x'} = \begin{cases} w_1^x, & x_i = x_i^h; \\ 0, & x_i \neq x_i^h. \end{cases}$$

**2. Згідно з кроком 1:**

$$2.1 \quad k = \frac{\sum_{i=1}^n w_i^{x'}}{\sum_{i=1}^n w_i^x}$$

**3. Завершити алгоритм.**

Таким чином, ми розраховуємо вектор ваг факторів поточного матчу  $\{W^{x'}\} = \{w_1^{x'}, \dots, w_n^{x'}\}$  та ваговий коефіцієнт  $k$  для подальшого виведення ймовірностей вдалого прийому, подачі або ситуації тай-брейку.

### 3.5.2 Фінальне прогнозування

На даному етапі прогнозування для подальшої роботи необхідно визначити значення оновлених ймовірностей  $p_j^{1'}$ ,  $p_j^{2'}$ ,  $p_j^{T'}$  - ймовірностей вдалої подачі, прийому, ситуації тай-брейку, базуючись на наших початкових ймовірностях  $p_j^1$ ,  $p_j^2$ ,  $p_j^T$  відповідно (згідно з пунктом 3.2).

Запишемо алгоритм розрахунку для нової  $p_j^{2'}$  ймовірності вдалого прийому гравця (аналогічно для  $p_j^{1'}$ ,  $p_j^{T'}$ ). Для цього використаємо початкову ймовірність  $p_j^2$  вдалого прийому, ваговий коефіцієнт  $k$  ( $0 \leq k \leq 1$ ) з пункту 3.5.1, проведемо нормалізацію отриманого значення та отримаємо нову ймовірність вдалого прийому гравця.

#### Алгоритм розрахунку оновлених ймовірностей

**Вхід:** початкова ймовірність  $p_j^2$  вдалого прийому, ваговий коефіцієнт  $k$ .

**Вихід:**  $p_j^{2'}$  - нова ймовірність вдалого прийому гравця.

#### 1. Для поточного матчу:

$$1.1 \quad p_j^{2'} = (1 + k)p_j^2$$

#### 2. Нормалізація отриманого значення ймовірності в діапазоні [0;1]:

$$2.1 \quad p_j^N = \frac{p_j^{2'} - \min(p_j^2)'}{\max(p_j^2)' - \min(p_j^2)'}$$

#### 3. Розрахувати оновлену ймовірність:

$$3.1 \quad p_j^{2'} = \frac{\sum_{i=1}^n w_i^{x'}}{\sum_{i=1}^n w_i^x}$$

Таким чином, ми розраховуємо оновлені ймовірності для вдалих подачі, прийому та ситуації тай-брейку.

Для розрахунку ймовірності виграшу гравцем матчу необхідно побудувати ланцюги Маркова для моделі загальної гри в теніс, враховуючи виведені нами раніше ймовірності, які тепер оновлено з урахуваннях статистичних даних гравця. Формула ймовірності виграшу гравцем матчу наведена у пункті 2.5.

### 3.6 Висновок до розділу

В цьому розділі був сформульований формальний опис алгоритмів розв'язання задачі. В кожному алгоритмі був виділений ряд елементарних операцій, що реалізуються окремо, та багатократно викликаються при виконанні основної частини алгоритму. Створення окремих тестів щодо кожної з цих елементарних операцій, декомпозиція тестів, дозволяє скласти процедури тестування, що мають властивість масштабованості.

На сьогоднішній день теніс являється одним із найпопулярніших видів спорту на планеті, що й зумовлює постійні модифікації та розробки сучасніших та точніших алгоритмів прогнозування результатів. Алгоритми базуються на різноманітних підходах, однак основні методи сьогодення – ланцюги Маркова та нейронні мережі.

Даний розділ показує етапи новоствореного алгоритму прогнозування результатів тенісного матчу. На основі наведеного алгоритму стало можливим комбінування ієрархічного стохастичного та машинного методів прогнозування результатів спортивних змагань.

Було сформульовано можливі подальші модифікації для кожного з підходів прогнозування.



## 4 ОПИС ПРОГРАМНОГО ПРОДУКТУ

Реалізацію програмного продукту було вирішено розроблювати у вигляді веб-додатку з інтуїтивно простим інтерфейсом користувача та можливістю практичного застосування розробленої методології та алгоритмів.

Веб-додаток повинен включати в себе функції виводу проаналізованих даних, можливість вибору конкретного матчу для аналізу та виведення результатів у зручній для користувача формі.

### 4.1 Структура компонент програмного забезпечення

Веб-застосунок отримує запит від клієнта і виконує обчислення, після цього формує веб-сторінку і відправляє її клієнтові мережею з використанням протоколу HTTP. Саме веб-застосунок може бути клієнтом інших служб, наприклад, бази даних або стороннього веб-застосунку, розташованого на іншому сервері. Прикладом веб-застосунку є система управління вмістом статей: безліч її учасників можуть брати участь у створенні мережевої енциклопедії, використовуючи для цього браузері своїх операційних систем (Microsoft Windows, GNU/Linux або будь-якої іншої операційної системи) без завантаження додаткових виконуваних модулів для роботи з базою даних статей.

Для більшої інтерактивності і продуктивності розроблений новий підхід до розробки веб-застосунків, названий AJAX, і який нині є стандартним де-факто. При використанні Аjax сторінки веб-застосунку здатні відправляти веб-запити до сервера у фоновому режимі, і не перезавантажуються цілком, а лише довантажують необхідні дані з сервера, що значно пришвидшує роботу і робить її зручнішою.

Архітектура клієнт-сервер є одним із архітектурних шаблонів програмного забезпечення та є домінуючою концепцією у створенні розподілених мережних застосунків і передбачає взаємодію та обмін даними між ними.

Вона передбачає такі основні компоненти:

- набір серверів, які надають інформацію або інші послуги програмам, які звертаються до них;

- набір клієнтів, які використовують сервіси, що надаються серверами;
- мережа, яка забезпечує взаємодію між клієнтами та серверами.

Сервери є незалежними один від одного. Клієнти також функціонують паралельно і незалежно один від одного. Немає жорсткої прив'язки клієнтів до серверів. Більш ніж типовою є ситуація, коли один сервер одночасно обробляє запити від різних клієнтів; з іншого боку, клієнт може звертатися то до одного сервера, то до іншого. Клієнти мають знати про доступні сервери, але можуть не мати жодного уявлення про існування інших клієнтів.

Діаграма компонентів програмної системи наведена на рисунку 4.1.

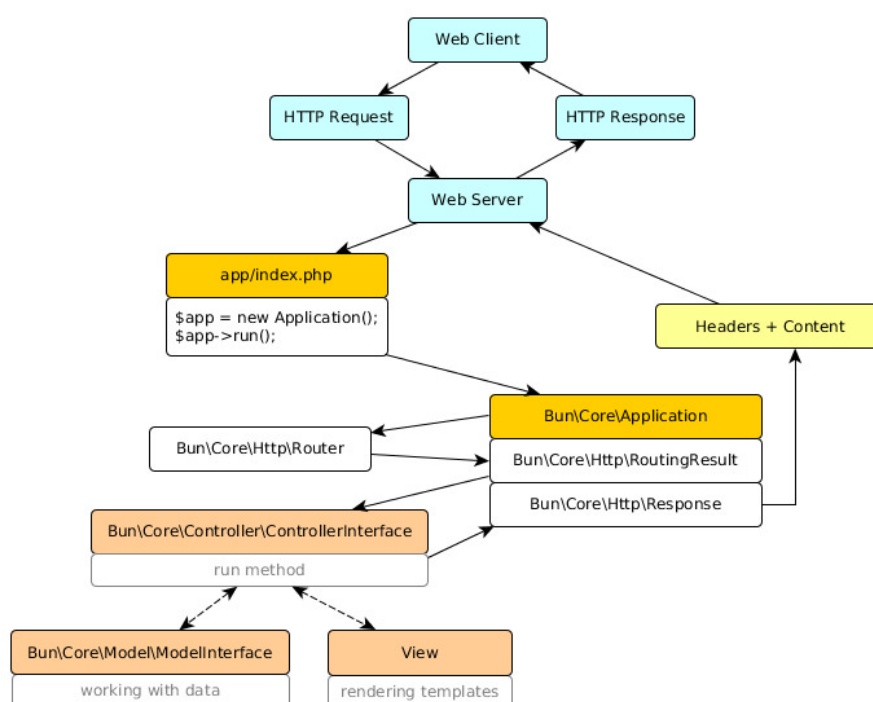


Рисунок 4.1 – Схема роботи нейронної мережі

## 4.2 Структура діаграми класів

Діаграма класів — статичне представлення структури моделі. Відображає статичні (декларативні) елементи, такі як: класи, типи даних, їх зміст та відношення. Діаграма класів, також, може містити позначення для пакетів та може містити позначення для вкладених пакетів. Також, діаграма класів може містити позначення деяких елементів поведінки, однак їх динаміка розкривається в інших типах діаграм. Діаграма класів (class diagram) служить для представлення статичної структури

моделі системи в термінології класів об'єктно-орієнтованого програмування. На цій діаграмі показують класи, інтерфейси, об'єкти й кооперації, а також їхні відносини.

Діаграма класів програмної системи наведена на рисунку 4.2.

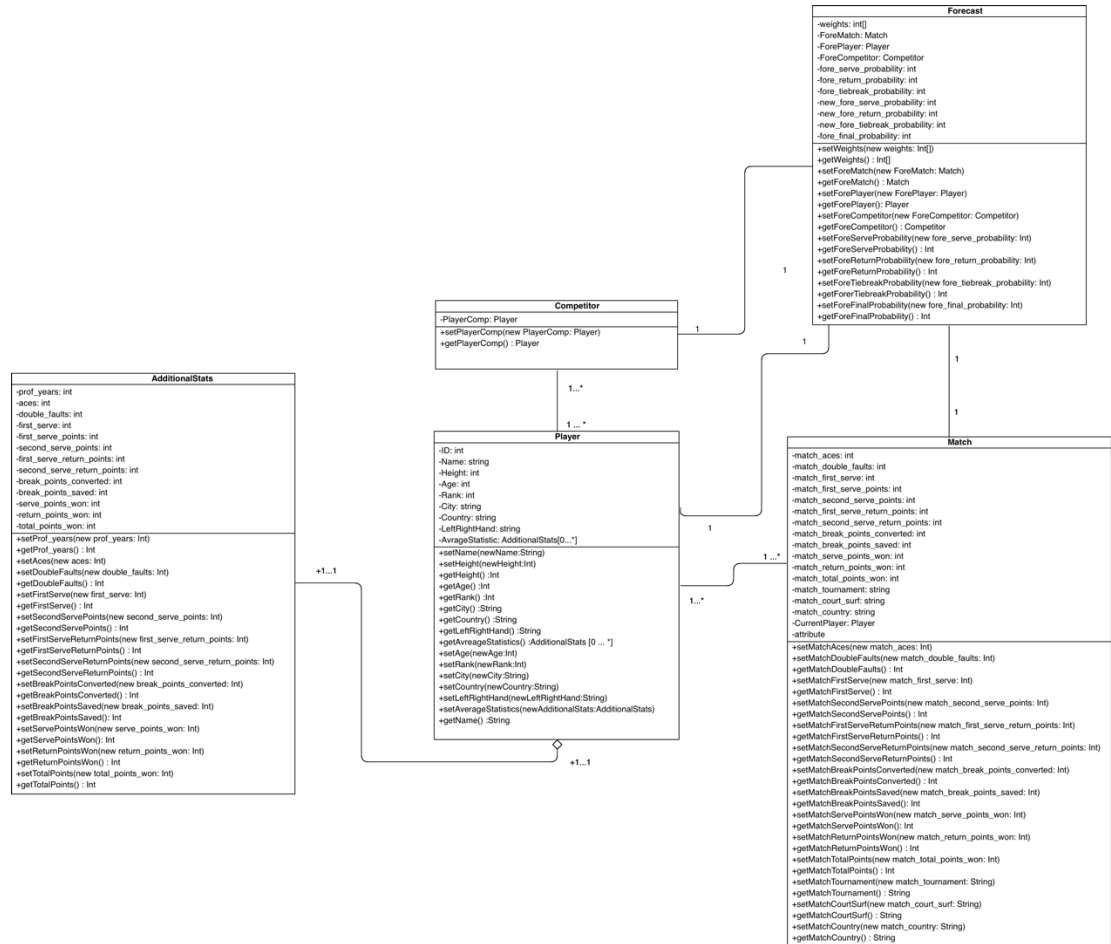


Рисунок 4.2 – Діаграма класів

Діаграма класів для побудови нейромережі програмної системи наведена на рисунку 4.3 [53].

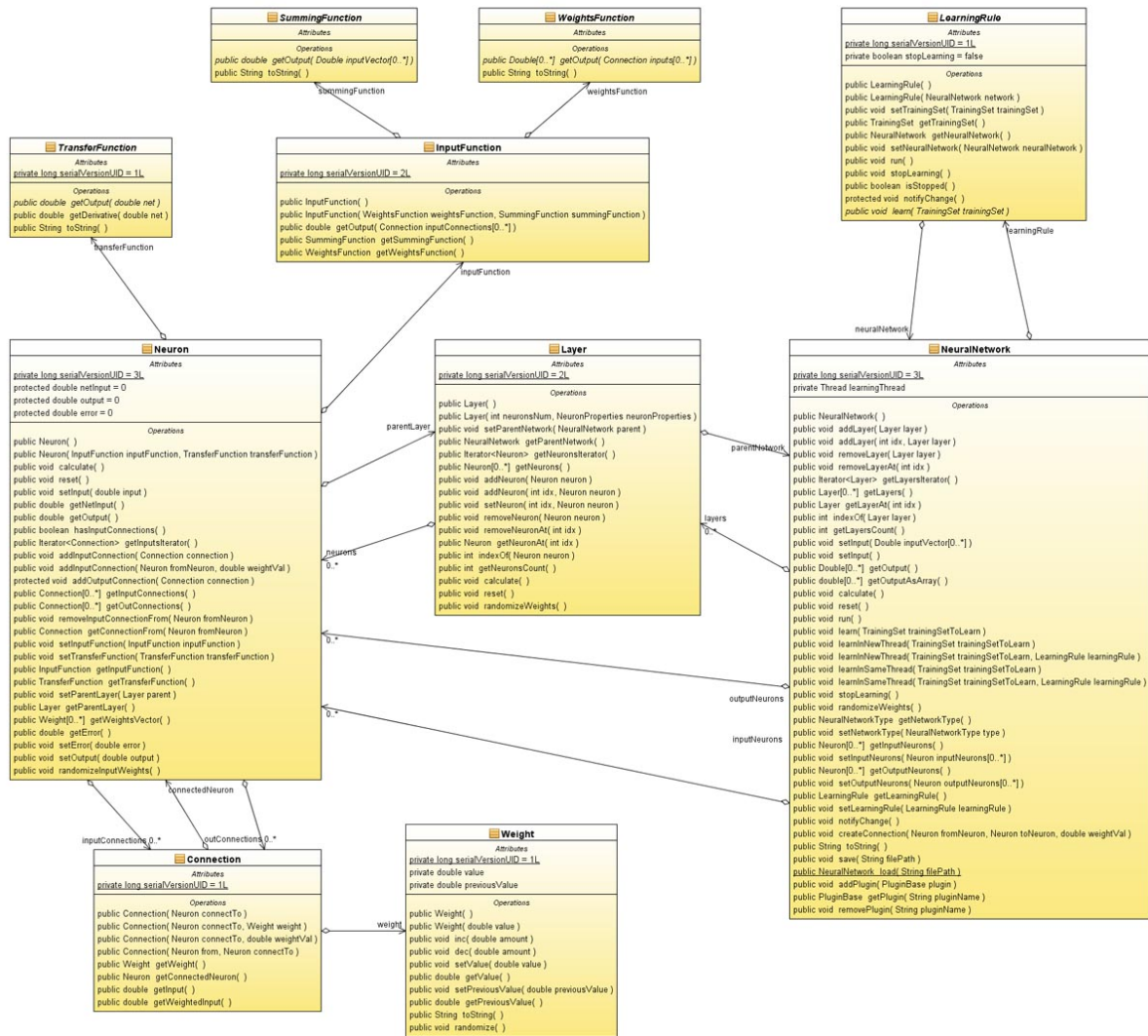


Рисунок 4.3 – Діаграма класів для побудови неймережі

### 4.3 Структура діаграми бази даних

За допомогою бази даних розміщуються дані про гравців, матчі, додаткові статистичні дані та результати прогнозування. На рисунку 4.4 наведено діаграму бази даних.

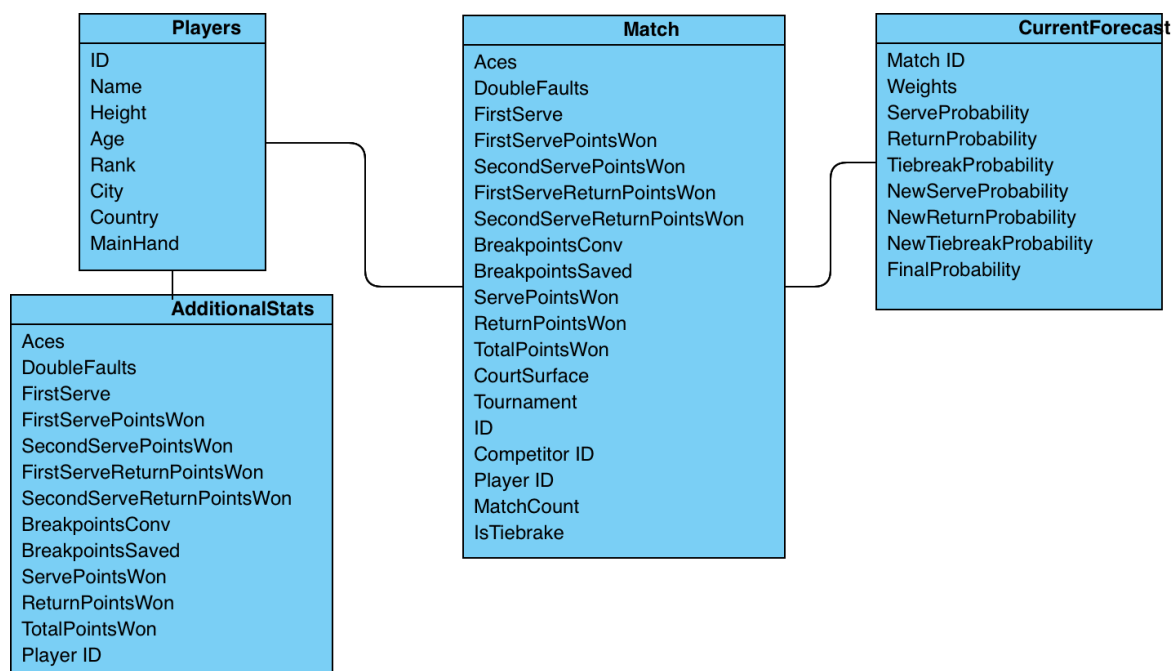


Рисунок 4.4 – Діаграма бази даних

#### 4.4 Структура use-case діаграми

Діаграма прецедентів є графом, що складається з множини акторів, прецедентів (варіантів використання) обмежених границею системи (прямокутник), асоціацій між акторами та прецедентами, відношень серед прецедентів, та відношень узагальнення між акторами. Діаграми прецедентів відображають елементи моделі варіантів використання. Структура use-case діаграми наведена на рисунку 4.5.



Рисунок 4.5 – Діаграма use-case

## 4.5 Керівництво користувача

Розпочати роботу з веб-додатком можна, відкривши його у веб-браузері. Верхні вкладки представляють головну навігацію додатком. Користувач може переглянути статистичні дані, натиснувши на вкладку “Статистичні дані”, розпочати прогнозування за допомогою ланцюгів Маркова у вкладці “Ланцюги Маркова”, розпочати прогнозування за допомогою нейронної мережі у вкладці “Нейронні мережі” та провести прогнозування за новою методологією, обравши вкладку “Прогнозування” або натиснувши на кнопку “Розпочати прогнозування”. На рис. 4.6 показано головну сторінку додатку.



Рисунок 4.6 – Головна сторінка додатку

При виборі одного з методів прогнозування користувач потрапляє на сторінку вибору даних. Тут користувачеві надається можливість обрати гравця, матч з яким хотілося б проаналізувати, натиснувши на його ім'я. За допомогою стрілки можливе повернення на головну сторінку. Випадаюче меню реалізує вибір вхідних даних – обирання гравця з бази програми чи завантаження власного CSV-файлу, що відповідає формату додатку. На рис. 4.7 показано головну сторінку додатку.

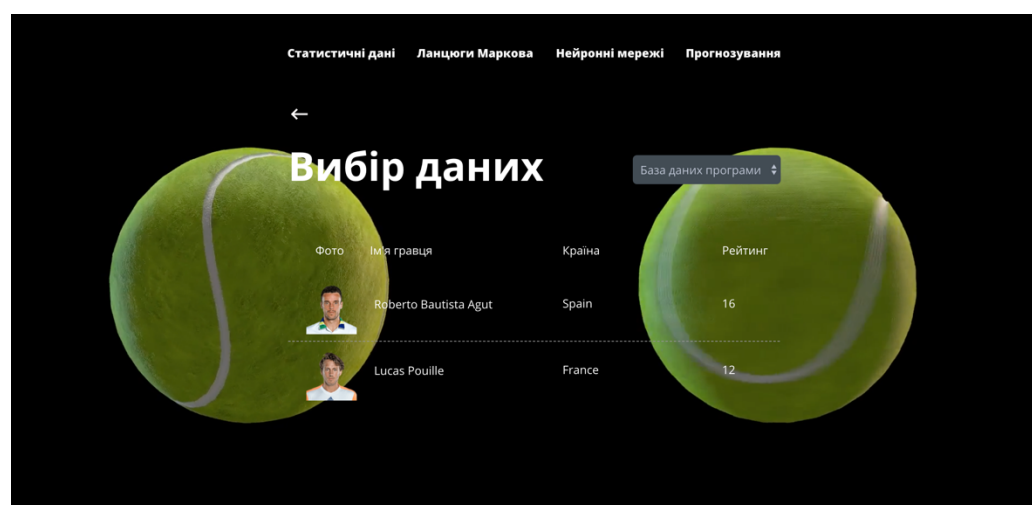


Рисунок 4.7 – Сторінка вибору даних

Після того, як користувач вибрав гравця, він потрапляє на сторінку вибору матчу. На ній можна бачити дані про гравця, обраного на попередньому екрані, а також усі матчі, які були ним зіграні та наявні у базі даних програми чи користувача. Вибір матчу здійснюється подібно до екрану вибору даних – натиском на ім'я суперника. На рис. 4.8 показано сторінку вибору матчу.

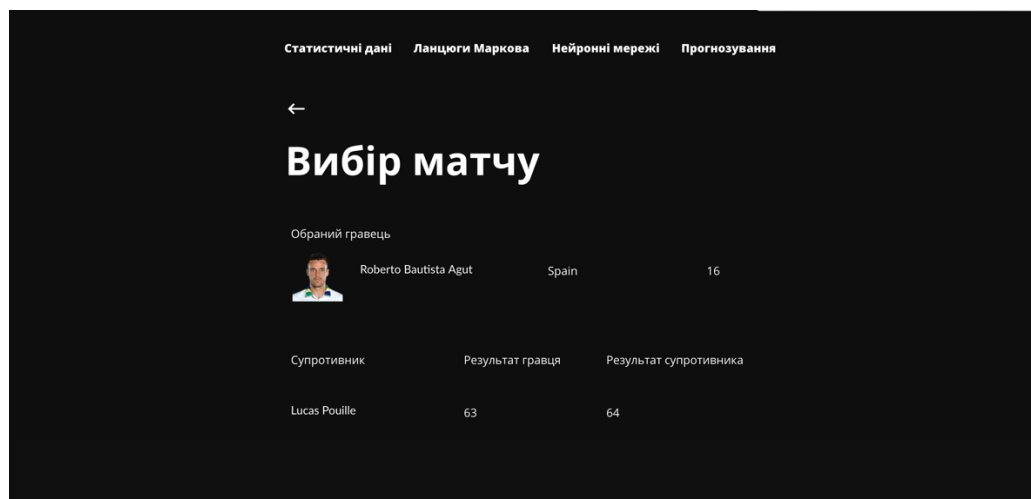


Рисунок 4.8 – Сторінка вибору матчу

Результати прогнозування виводяться на фінальному екрані. Користувач може розпочати прогнозування обраним методом ще раз або перейти на головну сторінку.

#### 4.6 Висновок до розділу

В даному розділі було описано програмне забезпечення, що використовувалося в ході створення програмного продукту, що реалізовує алгоритми прогнозування результатів спортивних змагань.

Керівництво користувача описує, як працювати з програмою та обирати дані для аналізу. Структуру програмного забезпечення наведено за допомогою діаграми класів, діаграми прецедентів, діаграми бази даних.



## **5 РОЗРОБКА СТАРТАП-ПРОЕКТУ**

### **5.1 Резюме стартапу**

#### **5.1.1 Назва проекту**

Дана стартап-розробка називається WingX-Ray і являє собою мобільний комплекс для цифрового радіологічного неруйнівного контролю в авіації.

#### **5.1.2 Ідея проекту**

Дослідження в авіаційній галузі. Практичне рішення за напрямком діагностики неруйнівного контролю – контролю властивостей та параметрів об’єкта, не руйнуючи його та при якому не повинна бути порушена здатність об’єкта до використання та експлуатації.

Суть технології в побудові з серії шарів (рентгенівських знімків) 3D-моделі об’єкта дослідження.

#### **5.1.3 Партнери проекту**

Партнери та учасники проекту:

- Згуровський М. З. – ректор Національного технічного університету України «Київський політехнічний інститут», доктор технічних наук, голова наглядової ради “Укроборонпром”, академік НАН України, Академії педагогічних наук України;
- Малюкова І. Г. – керівник Стартап Школи “Sikorsky Challenge”, кандидат технічних наук;
- Мірошніченко С. І. – професор, доктор технічних наук, академік АН Вищої освіти України, генеральний директор Науково-виробничого об’єднання “Телеоптик”;
- Новіков Ю. Л. – керівник проекту, кандидат технічних наук.

### 5.1.4 Технічні партнери проекту

Напрямами розвитку та технічними партнерами проекту є:

- а) авіація;
  - 1) Boeing [35];
  - 2) Progresstech [58];
  - 3) Антонов [3];
- б) військовий сектор;
  - 1) Укроборонпром [30];
  - 2) Артем [4];
- в) медицина;
- г) ветеринарія;
- д) геологічні розвідки.

## 5.2 Технологічний аудит ідеї проекту

### 5.2.1 Розвиток напрямку

Починаючи з першого рентгенографічного приймача фірми “SwissRay”, де зображення формувалося чотирма матрицями приборів із зарядовим зв’язком (ПЗЗ), конструкції з багатьма сенсорами ввійшли в техніку цифрових приймачів. Використання багатьох сенсорів (сенсор-SA), які формують часткові (парціальні) зображення, дозволили отримати відносно недорогі та оптимізовані по параметрам рентгенографічні, мамографічні і рентгеноскопічні приймачі [26].

Типова конструкція SA-приймача зображена на рисунку 5.1 [26].

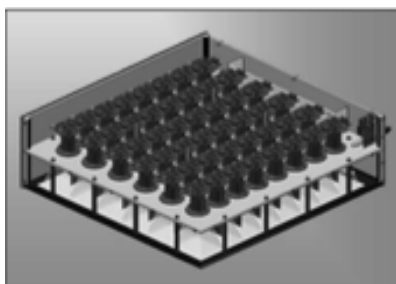


Рисунок 5.1 – Конструкція SA-приймача, пошарово, зверху вниз: поле сенсорів, об’єктиви, несуча панель, бленди, перетворювальний екран, дека

В той же час, доля SA-структур на світовому ринку відносно невелика – не більше 10%. Це зумовлено проявом в SA-структурах перешкод багатоканальності.

Приймачі в SA-структурах були розроблені в 90-х роках XX століття. До цього часу сформувалося три типи конструкцій рентгенівських приймачів.

Хронологічно першими були приймачі на надвеликих ПЗЗ-матрицях. На їх основі почався розвиток цифрових систем для загальної рентгенографії та систем дослідження органів грудної клітини. приймачі цього напрямку відрізняються простотою, однак, мають великі габарити, масу, достатньо низьку чутливість. Останнє зумовлено складністю виготовлення світлосильних об'єктивів для ПЗЗ-матриць великих розмірів. Роздільна здатність приймачів першого покоління була невисока та складала 2.5 ... 3 пар ліній на мм (п.л./мм).

На фоні виділених недоліків першого покоління приймачів на ПЗЗ-матрицях, проект, що передбачав створення приймачів у вигляді плоских панелей (Flat Panel Detectors – FPD) здавався ідеальним рішенням. Таким чином, плоскі панелі стали другим направленням в побудові цифрових рентгенівських приймачів.

Розробка пласких панелей вимагала інвестицій на рівні \$100 млн. доларів для кожного з проектів. Самі панелі мали роздільну здатність 3.5 ... 3.6 п.л./мм та високу квантову ефективність виявлення на низьких частотах. На жаль, обидва проекти мали серйозне внутрішнє протиріччя – невідповідність можливостей інтегральної технології (виготовлення сотень тисяч виробів у рік) та відносно малий об'єм замовлення. Це призвело до високої вартості панелі та низькому темпі розробок. До того ж, значні внутрішні шуми, наявність дефектних ділянок та схильність до деградації ускладнили експлуатацію панелей. Незважаючи на відмічені недоліки, цифрові приймачі на основі пласких панелей, починаючи з 2005 року, є лідерами за продажами на світовому ринку.

## **5.2.2 Порівняння з аналогами**

У 1997 році на виставці RSNA на стенді фірми Caresbuilt (США) були продемонстрована дослідна модель приймача для рентгенографії з полем 43x43 см та

вдвічі більшою (в порівнянні з пласкими панелями) роздільною здатністю – 7 п.л./мм. сукупна цифрова матриця приймача містила 6kx6k пікселів. Сумарна площа кремнію в приймачі складала 60x55 мм, а товщина приймача – близько 10 см.

Рекордні параметри приймача стали можливі завдяки застосуванню 99-матричної SA-конструкції, розробленої співробітниками компанії “Телеоптік”. В приймачі вперше було використане автокалібрування по ділянкам полів зору сенсорів, що перекриваються. В такій технології на зібраному зображенні присутні всі точки вихідного рентгенівського зображення, а границі парціальних зображень візуально не виявляються.

Рекордні характеристики демонстрував і мамографічний приймач фірми Bennett. Приймач складався з механічних щільно підігнаних волоконно-оптичних неоднорідних кабелів (фоконів), де великі торці примикали до перетворюючого екрану, а на менші торці були підклеєні матриці ПЗЗ. Наявність фоконів давала гарний оптичний зв'язок між перетворюючим екраном та матрицею ПЗЗ. Роздільна здатність складала 12 п. л./мм, що до сих пір є найкращим результатом для цифрових мамографічних приймачів.

Основними недоліками такої SA-конструкції були: втрата частини вихідного рентгенівського зображення на границях полів фоконів та відсутність перекриття полів матриць, що виключало автокалібрування. В результаті, через температурні дрейфи сенсорів, їх коефіцієнти передачі змінювалися. Як наслідок, на зображенні з'являлася перебивка у вигляді “шахової дошки”. Невдалість такої SA-конструкції породила представлення про складність та неефективність алгоритмів зклеювання парціальних зображень у багатосенсорних структурах.

На практиці можливість ефективного приглушення перебивок багатоканальності підтверджено успішною експлуатацією більш ніж 2 тисяч цифрових багатосенсорних приймачів з калібруванням протягом останніх 10 років.

### **5.2.3 Просторова модель та подавлення перебивок багатоканальності**

В 1996 році був запропонований підхід до побудови рентгено-телефізійної системи з високою роздільною здатністю. Він передбачає використання оптичного

тракту на основі комбінації “екран-масив” оптичних каналів, що складаються з об’єктивів та сенсорів. Відмінністю запропонованого підходу є процедура відновлення зображення при об’єднанні парціальних зображень з виходів оптичних каналів, що забезпечує відсутність втрати інформації на границі полів зору окремих сенсорів усунення геометричних та яскравісних спотворень зображень в кожному з каналів.

Коректне рішення задачі повного відновлення зображення на основі парціальних зображень передбачає знання кореляційних зв’язків на границях парціальних полів зору. в підході, що розглядається, інформація про кореляційні зв’язки забезпечується шляхом перекриття парціальних полів зору. При цьому, кожне з зображень може бути використане для калібровки системи. При перекритті парціальних зображень створюються умови для автокалібрування, тобто, для зклеювання по інформації в зонах перекриття поточного зображення.

Приймачі, що мають автокалібровку, безперечно, виграють в стійкості до змін зовнішніх умов та дрейфів параметрів сенсорів.

Навпаки, коли в таких випадках конструкція приймача не дозволяє безпосередньо змінити зв’язки між сигналами на границі парціальних зображень, а, тим паче, коли на границях частина пікселів повного зображення втрачається, ситуація стає принципово іншою. Необхідно по тестовим зображенням отримувати інформацію про кореляцію парціальних зображень. За допомогою такої інформації можливе відновлення повного зображення тільки при відсутності дрейфів параметрів сенсорів. В цілому, рішення задачі відновлення повного зображення по парціальним складовим в цьому випадку некоректне. При відсутності перекриття парціальних зображень існує ризик спотворення зображення в процесі склеювання.

#### **5.2.4 Склеювання зображення по калібрувальним тестам**

Модель спотворень зображень в SA-системах може бути представлена двома послідовними ланками. Перша описує просторові дисторсії парціальних зображень

$G_S\{X, Y, k\}$ , а друга – поелементні ефекти, що пов’язані зі зміною інтенсивності сигналу яскравості  $G_B\{X, Y, k, t, B\}$ . Загальний вираз для оператора спотворення :

$$G\{X, Y, k\} = G_S\{X, Y, k\}G_B\{X, Y, k, t, B\}, \quad (6.1)$$

де  $X, Y$  – просторові координати;

$t$  – час;

$k$  – номер оптичного каналу;

$B$  - інтенсивність сигналу яскравості.

Задача відновлення зображення полягає в знаходженні зворотного перетворення, що описує усунення спотворення (6.1), що вносяться в процесі формування зображення:

$$B'\{X, Y, t\} = G^{-1}\{X, Y, k, t, B(X, Y, t, k)\}, \quad (6.2)$$

де  $B'\{X, Y, t\}$  - відновлене зображення.

При цьому, ціллю відновлення є отримання зображення, придатного для наступного аналізу оператором.

Спотворення зображень можна розділити на детерміновані в часі та ті, що мають залежність від часової координати. До першого типу відносяться геометричні, а також поелементні яскравісні спотворення, що описують світлові характеристики окремих пікселів оптичного каналу.

Просторові (геометричні) спотворення зумовлені наступними факторами: геометричною дисторсією об’єктива, неперпендикулярністю оптичної осі об’єктива та площини сенсора, геометричною неоднорідністю площини перетворювального екрану (джерела світла). Для знаходження зворотного оператора  $G_S^{-1}\{X, Y, k\}$ . Було запропоновано використовувати метод апріорного моделювання на основі аналізу спотворень тест-об’єктів. В якості тест-об’єкта була використана лінійна сітка, причому кожен лінійний відрізок всередині тесту спостерігається як мінімум двома сусідніми сенсорами.

Схема спотворення зображення клітини сітки зображена на рисунку 5.2 [26].

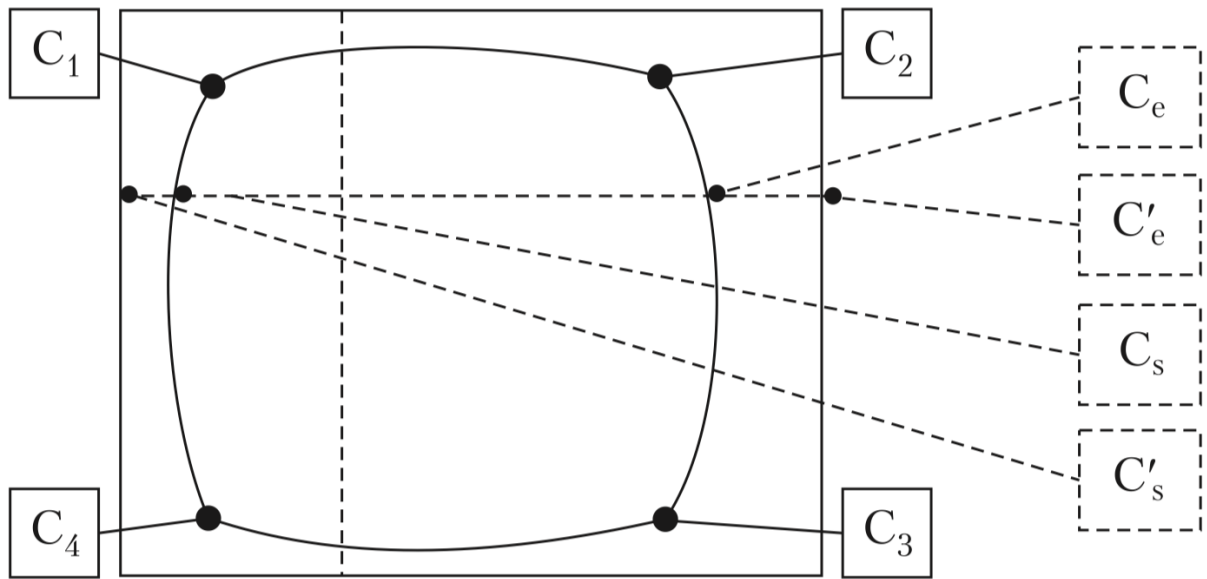


Рисунок 5.2 – Спотворення зображення клітини сітки

Процедура корекції полягає у відновленні вихідної прямокутної форми сітки.

Перетворення координат зображень (корекція) по горизонталі виконується відповідно зі співвідношенням:

$$C' = C_S' + C \frac{C_e' - C_S'}{C_e - C_S}, \quad (6.3)$$

де  $C$  - координати  $X(Y)$  пікселів на вхідному зображенні;

$C'$  - координати пікселя вихідного скоректованого зображення.

Аналогічно виконується корекція по вертикалі.

Якість відновлення геометричних спотворень зображення визначається точністю визначення координат елементів тест-об'єкта (ліній або точок сітки). Для підвищення точності відновлення оператора геометричних спотворень виконується поліноміальна апроксимація координат центрів ліній або точок сітки, обчислених за алгоритмом пошуку центра мас.

Другий тип детермінованих в часі спотворень зображень  $G_B\{X, Y, k, t, B\}$  пов'язаний з їх яскравістю. Він зумовлений неідентичністю коефіцієнтів пропуску об'єктів по полю, світлових характеристик сенсорів та постійною складовою коефіцієнтів посилення сигналу з виходу матриць.

Задача відновлення виду оператора  $G_B\{X, Y, k, t, B\}$  вирішується на основі аналізу серій зображень чистого поля, при цьому тестові зображення являють собою

зображення, отримані для різних значень рентгенівських доз при відсутності об'єктів в межах поля огляду.

Корекція яскравісних спотворень зображення виконується у відповідності з виразом:

$$B_{\text{вих}}(X, Y) = \left( B_{\text{вх}}(X, Y) - C_{F1}(X, Y, n) \right) C_{F2}(X, Y, n) + B_{\text{оп}}(n), \quad (6.4)$$

де  $C_{F1}$  - відступ  $n$ -го тесту;

$C_{F2}$  - нахил лінії  $n$ -ого тесту;

$n$  - номер тесту;

$B_{\text{оп}}(n)$  - значення опорної яскравості для корекції  $n$ -го тесту;

$B_{\text{вх}}(X, Y)$  - вхідна яскравість;

$B_{\text{вих}}(X, Y)$  - вихідна яскравість.

Для поля сенсорів процедура корекції реалізується наступним чином: за максимальною яскравістю на зображеннях тест-об'єктів обирається опорна точка. Задачею виконання перетворення яскравості є вирівнювання тестових зображень до яскравості опорної точки. Вхідна яскравість зображення кожного сенсору перераховується за формулою (6.4), утворюючи для кожного тесту рівнояскраве поле.

### 5.2.5 Автокалібрування

Описана в розділі 5.2.4 процедура яскравісної корекції зображень за калібрувальними тестами, є коректною за умови, що неідентичність пропускання об'єктів по полю, світлові характеристики сенсорів та коефіцієнти посилення сигналу з виходу матриць не змінюються з часом. На жаль, на практиці це не виконується. В результаті яскравісна корекція з часом, а також при зміні температури сенсора погіршується. Доводиться заново формувати калібрувальні дані або використовувати більш ефективну обробку рентгенівських зображень (автокалібрування) [26].

Вид оператора, що характеризує часову мінливість параметрів сигналів оптичних каналів, для поточного зображення можна відновити на основі методу апостеріорного моделювання при аналізі параметрів зображень сусідніх оптичних



каналів в областях, що перекриваються. Параметри корекції спотворень обчислюються шляхом оцінок статистичних характеристик зображень в зонах перекриття.

Наявність зон перекриття дозволяє вирішити задачу відновлення єдиного зображення з відновлених парціальних зображень на виході оптичних каналів та усунення розривів в точках стикування парціальних зображень. Для цих цілей використовується локальна просторова фільтрація в межах областей, що перекриваються, при цьому враховуються особливості зорової системи оператора, який виконує наступний аналіз зображення. Яскравість в одних і тих самих точках в зонах перекриття, що спостерігається в різних каналах – прирівнюються, в результаті зображень в сусідніх каналах – підтягуються.

У відповідності з записаною вище процедурою будується обчислювальний процес. При цьому помилка в відновленні геометричних спотворень не перевищує  $\frac{1}{4}$  пікселя, а похибка у відновленні яскравості не перевищує 0.1-0.5%, що робить границі зображень в сусідніх каналах візуально не спостережуваними. Обробка виконується з темпом у 5 мільйони пікселів у секунду. Час виконання корекції для одного рентгенівського зображення на серійному персональному комп'ютері складає 2-3 секунди.

### **5.3 Застосування результатів досліджень, проведених в дисертації, для побудови схеми розміщення сенсорів**

Для аналізу щільності перебивок необхідно враховувати наступні особливості результатів роботи сенсорів:

- кількість гарячих пікселів (при високій, низькій яскравості та у темряві);
- кількість битих пікселів (при високій, низькій яскравості та у темряві);
- рівень шуму на довгих витримках;
- рівень шуму при використанні різної оптики.

Перелічені характеристики впливають на класифікацію сенсорів за впливом на спотворення зображень та дозволяють провести власний аналіз постачальних

матеріалів, а також є етапом попередньої обробки вхідних даних для побудови схеми розміщення сенсорів в SA-приймачі.

На основі класифікаційних даних та загального плану розміщення сенсорів в SA-приймачі можливо змодельовати схему розташування сенсорів на несучій панелі шляхом переставляння та обертання сенсорів таким чином, щоб мінімізувати сумарні перебивки областей, що склеюються.

Задача побудови схеми розташування сенсорів вирішується на основі досліджень, проведених в даній дисертаційній роботі, шляхом проектування та аналізу варіантів розташування сенсорів.

## **5.4 Аналіз ринкових можливостей запуску стартап-проекту**

### **5.4.1 Поточний стан ринку продукту**

Проблематика ринку NDT (Nondestructive testing – неруйнуючий контроль) є наступною:

- неможливість якісної діагностики вузлів з композитних матеріалів та збірних приладів;
- неможливість якісної мобільної діагностики ключових вузлів повітряних суден в зонах конфліктів та стихійних лих;
- неможливість проведення динамічного контролю пристроїв в недоступних частинах фюзеляжу;
- високі утримання по доставці з повітряних суден на спеціалізовані точки проходження різних видів планового контролю та техобслуговування.

### **5.4.2 Рішення**

Адаптувати для ринку NDT в авіації існуючий прототип, що дозволяє здійснювати:

- глибинне пошарове дослідження – томосинтез;

- дослідження вузлів з композитних матеріалів та складних механізмів у збірці;
- динамічне дослідження в реальному часі;
- отримання результатів в цифровому форматі з високою чіткістю.

Існуючий прототип повинен володіти наступними експлуатаційними перевагами:

- мобільність;
- можливість використання будь-яких джерел енергії;
- працеспроможність в складних кліматичних умовах;
- стійкість до зовнішніх факторів: удари, волога;
- модульна конструкція.

## **5.5 Розроблення ринкової стратегії**

### **5.5.1 Історія впровадження**

У 2004 році об'єм продажів приймачів по SA-технології склав 4%, а в 2012 році - близько 5% світового об'єму виробництва всіх видів цифрових приймачів [26].

Популярність багатосенсорних приймачів з автокалібровкою з кожним роком зростає в зв'язку з високими техніко-економічними та експлуатаційними характеристиками. На даному етапі однією з останніх розробок SA-приймача є SA-приймач "ІОНА-Р-4343", що має роздільну здатність 4.6 п.л./мм (найвищу з аналогів) та в середньому втричі меншу товщину – 22 см та масу – менше 25 кг. Завдяки цьому даний приймач має гарну сумісність зі штативами практично будь-яких рентгено-діагностичних комплексів.

### **5.5.2 Ринкова стратегія**

Загальний об'єм ринку NDT в авіації – \$608.25 млн.

Очікується щорічний ріст у розмірі 9.15%.

Об'єм ринку пристроїв NDT – \$243.3 млн.

Найбільш перспективні сегменти ринку:

- Латинська Америка - \$15.8 млн.;
- Середній Схід та Африка - \$20.7 млн.

Оціночна потреба – 3000 пристроїв на рік.

### 5.5.3 Огляд конкурентів

Таблиця порівняння ключових конкурентів з розробленим рішенням наведена таблиці 5.1.

Таблиця 5.1 – Порівняння ключових конкурентів

Характеристики	WingX-Ray	GE DXR250U-W	VIDISCO FlashX Pro
Діагностика пошкоджень, що є невидимими при рентгенографії (наприклад, тріщини під кутом)	+	-	-
Виявлення глибини пошкоджень	+	-	-

Характеристики	WingX-Ray	GE DXR250U-W	VIDISCO FlashX Pro
Динамічне дослідження в режимі реального часу	+	-	-
Об'єм даних, що збирається для наступного аналізу	1 Gb	0.02 Gb	0.02 Gb

#### 5.5.4 Технологічні переваги

Технологічними перевагами даного рішення є:

- модульна конструкція;
- висока ремонтоспроможність;
- висока роздільна здатність: до 100 мікрон;
- широкий діапазон напруг: 20-270 кВ;
- широкий спектр розмірів полів огляду: від 24x30 см до 43x60 см.

#### 5.6 Прототипи продукту

На рисунку 5.3 зображена установка, що використовується для томосинтезу[26].

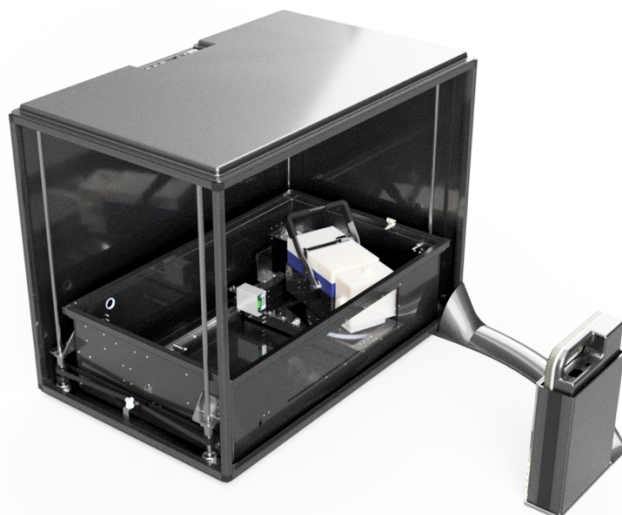


Рисунок 5.3 – Приклад установки, що використовується для томосинтезу

На рисунках 5.4-5.5 зображений процес виявлення пошкоджень лопасті пропелера за допомогою радіографії [26].



Рисунок 5.4 – Лопасть пропелера, що піддається аналізу

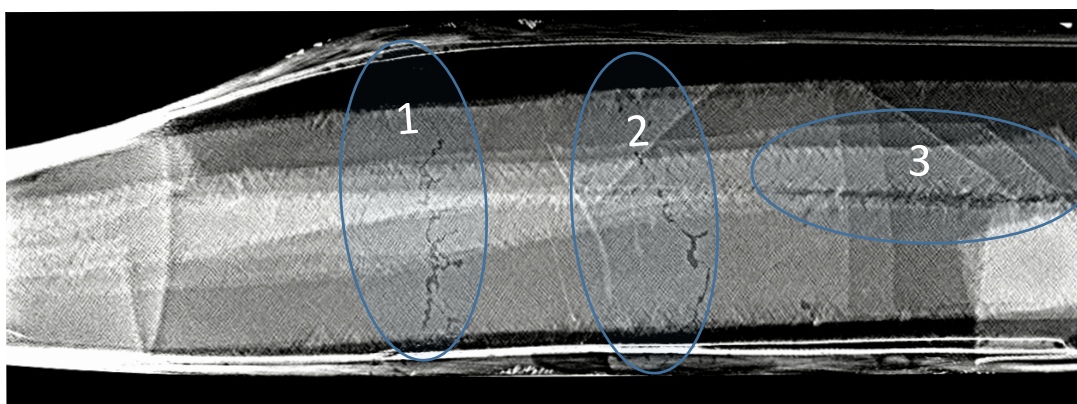


Рисунок 5.5 – Результат радіографічного аналізу

При проведенні радіографічного аналізу були досягнуті такі результати:

- роздільна здатність: 100x100 мікрон;
- час обробки зображення: 5 с.

На рисунку 5.6 зображено результат досліджень тієї ж лопасті пропелера за допомогою томосинтезу [26].

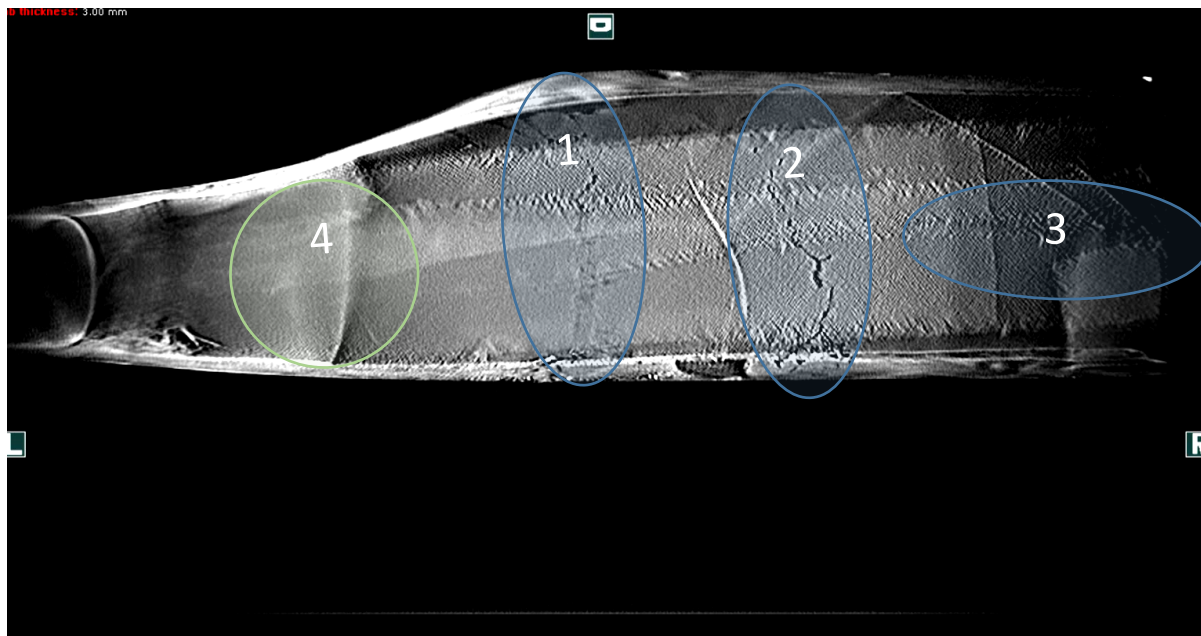


Рисунок 5.6 – Результат аналізу за допомогою томосинтезу

При проведенні аналізу за допомогою томосинтезу були досягнуті такі результати:

- кількість проєкцій: 70;
- час обробки однієї проєкції: 2.5 с;
- час створення 3-D моделі: 1-2 хв;
- об'ємна роздільна здатність: 250x250x1000 мікрон.

## 5.7 Побудова бізнес-моделі

### 5.7.1 Бізнес-модель

Модель: виробництво та продаж обладнання.

Канали збуту: канал продажів інвестора, мережі дистриб'юторів та дилерів.

Потенційні клієнти: аеропорти, аеродроми, інжинірингові компанії з обслуговування та техогляду, авіакомпанії.

### 5.7.2 Прогноз розвитку бізнесу

Ринкова вартість комплексу - \$115 тис.

Прогноз продажів:

- 1-й рік – створення трьох приладів для введення у дослідну експлуатацію;
- 2-й рік – реалізація 10 приладів;
- 3-й рік та далі – 10-15 та більше приладів у рік.

Ринки:

- в перші роки – Європа, Африка, Південна Америка;
- в подальшому – США, Канада, Азія, тихоокеанський регіон.

Досягнення точки беззбитковості: 2 роки.

Термін окупності проекту: 4 роки.

### 5.7.3 Досягнення

Налаштовано виробництво та продаж мобільних стаціонарних приладів для томосинтезу в сферах медицини та ветеринарії (3-4%).

## 5.8 Висновок до розділу

Запропонований продукт WingX-Ray являє собою мобільний комплекс для здійснення цифрового радіологічного неруйнівного контролю. Даний продукт може бути використаний для широкого кола потреб у сферах медицини, ветеринарії, геології, авіації.

Було описано технологічні особливості роботи SA-приймачів та описано практичне застосування методів прогнозування та використання результатів досліджень для контролю комплектуючих та побудови схем розміщення сенсорів на несучій панелі.

Проект зайняв 2 місце на VI Конкурсі Стартапів “Sikorsky Challenge” 11-12 жовтня 2017 року та був представлений на наглядній раді “Укроборонпрому” у березні 2018 року.



Також було проведено бізнес-аналіз стартапу, означено його роль на ринку подібних технологій та побудовано прогноз розширення ринку збуту.

В додатку А знаходиться копія сертифікату, що засвідчує проходження навчання у Стартап Школі “Sikorsky Challenge”.

## 6 ЕКСПЕРИМЕНТИ

### 6.1 Застосування модифікованого методу Маркова

В якості прикладу проведемо модифікацію наведеного раніше способу побудови ланцюга Маркова для конкретного тенісного між Рафаелем Надалем з Маріном Чиличем на Australian Open 2018 року.

Знайдемо умовні ймовірності отримання очок кожним із гравців та ймовірність виграшу гейму Чиличем за умови подачі та прийому, побудувавши таблиці умовних ймовірностей для обох випадків та візуалізуємо їх марківськими моделями гейму.

Візьмемо за основу гейм і вважатимемо рахунок простором станів. Побудуємо модель для гравця  $A$  за умови, що не він подавав (приймав) – оберемо деяке  $s$ , що означатиме ймовірність виграшу очка гравцем  $A$  за умови, що подає (приймає) не він, а суперник  $B$ . Ймовірність виграшу очка суперником  $B$  позначимо як  $l$ .

Для нашого прикладу, скориставшись статистичними даними матчу, встановлюємо, що ймовірність виграшу очка при подачі Чиличем – це  $s = 0,77$ , тоді ймовірність виграшу очка Надалем – це  $l = 1 - s = 0,23$ .

При прийомі Чиличем – це  $s = 0,41$ , тоді ймовірність виграшу очка Надалем – це  $l = 1 - s = 0,59$ .

Побудуємо марківський процес із дискретним часом, кінцевим станом та матрицею переходів  $M$  і початковим станом  $I_0$ . Множина станів цього ланцюга являється поглинаючою (або замкненою), так як включає в себе 2 стани, з яких не може бути переходу в наступний стан – це стан виграшу гравця  $A$  та стан виграшу гравця  $B$ . Оскільки ми модифікуємо ситуацію гри гейму, то матриця  $M$  набуває вигляду:

$$M = \begin{bmatrix} m_{11} & \cdots & m_{117} \\ \vdots & \ddots & \vdots \\ m_{171} & \cdots & m_{1717} \end{bmatrix}$$

Це зумовлено тим, що при моделюванні марківського ланцюга для гейму гри може бути всього 17 станів системи. Запишемо загальний вигляд блочної структури даної матриці:

$$M = \begin{bmatrix} E & 0 \\ R & W \end{bmatrix}$$

Тут  $E$  - одинична підматриця, порядок якої збігається з числом поглинаючих станів;  $W$  - квадратна підматриця ймовірностей переходів на множині неповоротних станів;  $R$  - прямокутна підматриця переходів з неповоротних станів в поглинаючі;  $0$  - нульова підматриця. Якщо число загальних станів – 17, з них 2 – поглинаючі стани, тоді підматриці мають наступний порядок:

$$M_{17 \times 17}, E_{15 \times 15}, W_{2 \times 2}, R_{2 \times 15}, 0_{15 \times 2}$$

Результати обчислень умовних ймовірностей отримання очка при подачі наведені у таблиці 6.1:

Таблиця 6.1. Таблиця умовних ймовірностей отримання очка при подачі

Чилич		Надаль				
		0	15	30	40	Гейм
	0	0.962890674	0.896661648	0.737136081	0.419136599	0
	15	0.98267337	0.944312142	0.832122939	0.544333246	0
	30	0.994131919	0.977823203	0.918086095	0.706926293	0
	40	0.999003354	0.995666754	0.981159802	0.918086095	
	Гейм	1	1	1		

Знайдемо умовні ймовірності отримання очок кожним із гравців та ймовірність виграшу гейма Чиличем за умови, що він приймає подачу. Для цього побудуємо таблицю умовних ймовірностей та візуалізуємо її марківською моделлю гейму.

Результати обчислень умовних ймовірностей отримання очка при прийомі наведені у таблиці 6.2:

Таблиця 6.2. Таблиця умовних ймовірностей отримання очка при подачі

Чилич		Надаль				
		0	15	30	40	Гейм
	0	0.285501053	0.173073318	0.081225575	0.022444053	0
	15	0.447287306	0.30524446	0.165813619	0.054741592	0
	30	0.651690425	0.505888841	0.325648973	0.133516079	0
	40	0.86150246	0.765258408	0.602132894	0.325648973	
	Гейм	1	1	1		

Ці данні показують, що Чилич зможе виграти гейм при прийомі подачі з ймовірністю 28%.

Візуалізацію ланцюга за умови, що Чилич приймає подачу наведено на рисунку 6.1:

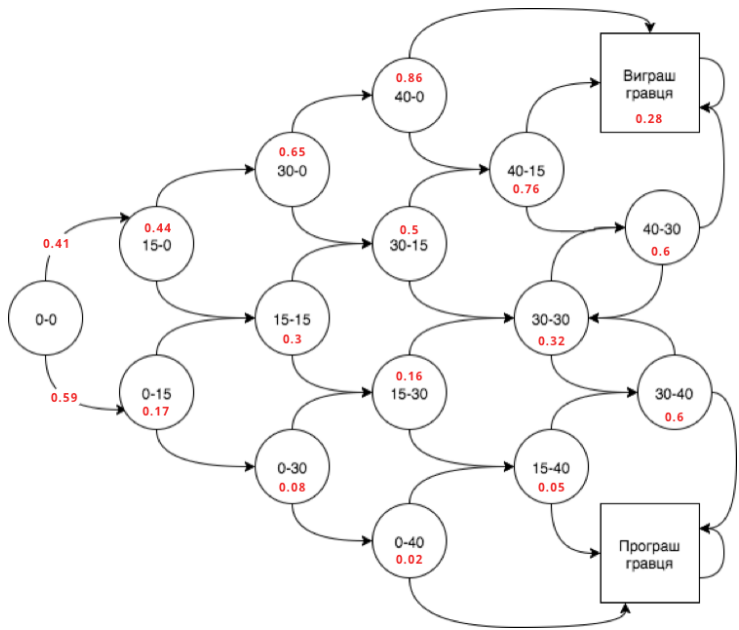


Рисунок 6.1 - Марківська модель гейму в матчі за умови, що приймає подачу  
Марин Чилич

Візуалізацію ланцюга за умови, що Чилич подає наведено на рисунку 6.2:

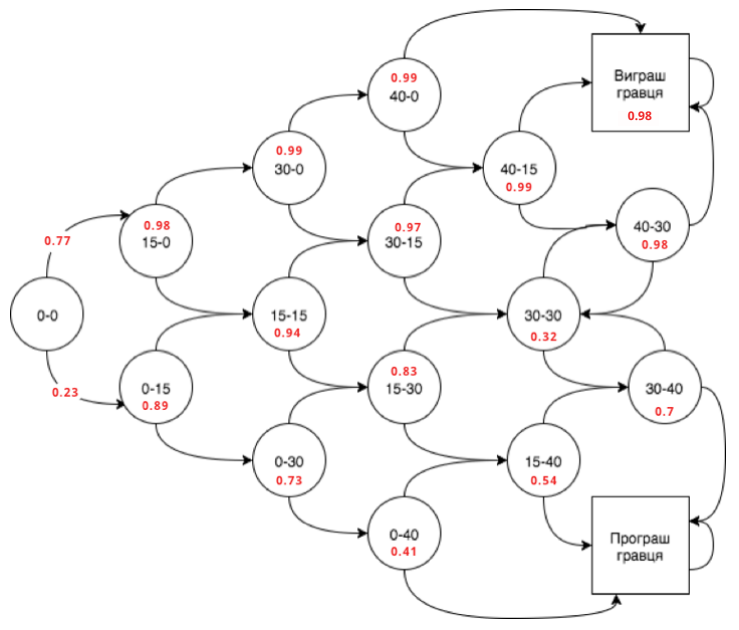


Рисунок 6.2 - Марківська модель гейму в матчі за умови, що подає Марин  
Чилич

Порівнявши отримані данні з даними статистики матчу можемо вважати, що даний спосіб аналізування тенісного гейму може бути використаний для практичного застосування та давати результати, що цілком узгоджуються з реальними даними.

## 6.2 Дослідження ефективності алгоритму прогнозування результатів спортивних змагань

Даний експеримент досліджує ефективність роботи створеного в ході роботи алгоритму, що включає в себе комбінування стохастичного та машинного методів прогнозування результатів тенісних змагань на прикладі гри в теніс.

Загальними даними для прогнозування стали статистичні показники по 374 матчах. Вибір 16 основних факторів впливу на результати прогнозування був заснований на результатах дослідження Ма, Ліу і Тана [9], які за допомогою нейромережевого підходу, заснованого на логістичній регресії визначають залежну змінну – ймовірність виграшу гравця у матчі за даних при цьому умов.

Наведемо конкретні дані результатів аналізу для матчу Роджера Федерера з Танасі Кокінакісом. Перш за все, необхідно знати середньостатистичні дані гравця для початку прогнозу. Дані Федерера наведено у таблицях 6.3, 6.4.

Таблиця 6.3 – Дані про гравця Роджера Федерера

<i>Гравець</i>	<i>Вік</i>	<i>Зріст, см</i>	<i>Головна рука</i>	<i>Ранг</i>	<i>Місто</i>	<i>Країна</i>
Роджер Федерер	36	185	Права	1	Базель	Швейцарія

Таблиця 6.4 – Загальні статистичні дані Роджера Федерера

<i>Дані Роджера Федерера</i>			
Ейси	10463	Зіграно подач	17221
Подвійні помилки	2555	Зіграно прийомів	16832
Перша подача	62%	Виграно прийомів загалом	51%
Виграно очок при першій подачі	77%	Виграно очок при першому прийомі	33%
Виграно очок при другій подачі	57%	Виграно очок при другому прийомі	51%
Брейк-поінти	5919	Конвертовані брейк-поінти	41%
Брейк-поінти збережені	67%	Загальна кількість виграних очок	54%

При побудові початкових ланцюгів Маркова і роботи з нейронною мережею було визначено ваги факторів для тенісиста при вхідних даних. Діапазон значень факторів –  $[0;1]$ . Таблиця визначених впливових факторів та їх ваг наведена у таблиці 6.5.

Таблиця 6.5 – Таблиця впливових факторів та їх ваг

Глиняне покриття	Прийом подачі	Ейси
0.21	0.303	0.15

Отже, з даних, наведених щодо вагомих факторів можна рекомендувати гравцю більше тренуватися на глиняному покритті, відпрацьовувати прийом подач та ейси перед матчем, що має відбутися.

Під час перерозрахунку ймовірностей було вираховано наступні ймовірності, отримані у таблиці 6.6.

Таблиця 6.6 – Ймовірність виграшу у геймі (початкова та оновлена)

	<i>Подача</i>	<i>Прийом</i>	<i>Тай-брейк</i>
<i>Ймовірність виграшу у геймі (до обчислень)</i>	0.77	0.33	0.45
<i>Оновлена ймовірність виграшу</i>	0.73	0.38	0.44

Після проведення фінальних розрахунків та побудови фінальної моделі було визначено, що Роджер Федерер виграє у даному поєдинку із ймовірністю 63%.

Даний матч вже відбувся і тому ми можемо говорити про його реальний результат. Перемогу у матчі здобув Кокінакіс із невеликим розривом. Такий результат може бути науково пояснений, так як базуючись на даних досліджень результатів тенісних турнірів було доведено, що у двобої між молодим та більш досвідченим професійним гравцем із багаторічним стажем з більшою вірогідністю переможе молодший. Передусім, значний вплив може мати непристосованість професійного спортсмена до гри та енергійності більш молодого гравця.

Засновуючись на даних про відсоток коректних прогнозів шляхом стандартного методу ланцюгів Маркова та нейронних мереж, було виведено дані про коректність роботи алгоритму, наведені на рисунку 6.2.

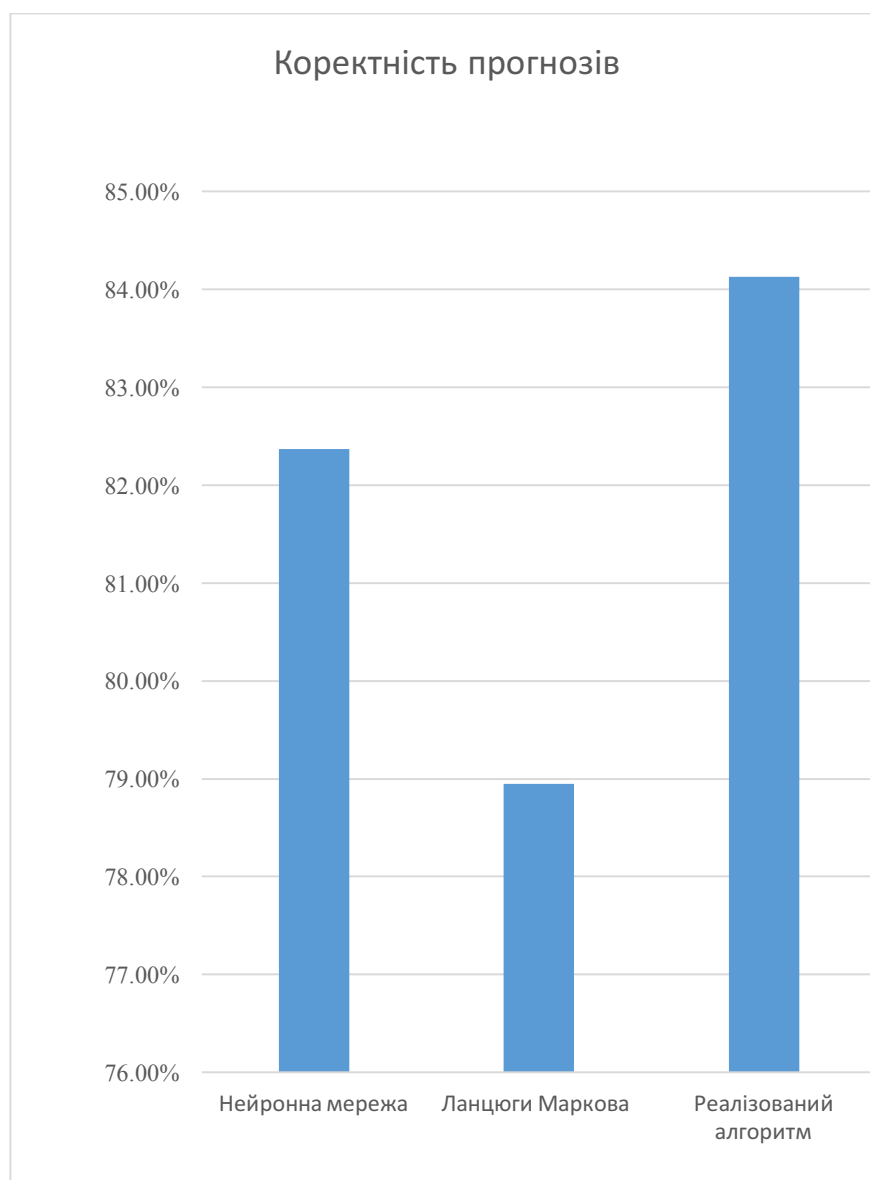


Рисунок 6.2 – Коректність прогнозів у відсотковому співвідношенні



## ПЕРЕЛІК ПОСИЛАНЬ

1. Айвазян С. А. Прикладная статистика: классификация и снижение размерности. [Текст] / Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. // М.: Финансы и статистика, 1989.
2. Акулич И. Л. Математическое программирование в примерах и задачах. [Текст] / Акулич И. Л. // М.: Высшая школа, 1986. — с. 298-310.
3. Антонов [Электронный ресурс] // Режим доступа: <http://www.antonov.com>
4. Артем [Электронный ресурс] // Режим доступа: <http://www.artem.ua/ru/>
5. Б. Фоменко. Теннис. Справочник. [Текст] / Б. Фоменко // Москва : «Физкультура и спорт», 1981. — 136 с.
6. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. [Текст] / Вапник В. Н. // М.: Наука, 1979.
7. Г. Соколов. Модель игры в теннис - Марковская цепь [Электронный ресурс] // Режим доступа: <http://www.sportradar.ru/article/math-for-sports-tennis.html>
8. Далингер В.А. Обучающимся о применении математики в спортивной сфере деятельности человека [Текст] / Далингер В.А., Федоров В.П. // Международный журнал прикладных и фундаментальных исследований. — 2014. — № 3 (часть 1) — 105-109 с.
9. Енциклопедія українознавства : Словникова частина [Текст] // Наукове товариство імені Шевченка ; Париж ; Нью-Йорк : Молоде життя ; Львів ; Київ : Глобус, 1955—2003.
10. Журавлев Ю. И. «Распознавание». Математические методы. Программная система. Практические применения. [Текст] / Журавлев Ю. И., Рязанов В. В., Сенько О. В. // М.: Фазис, 2006.

11. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. [Текст] / Загоруйко Н. Г. // Новосибирск: ИМ СО РАН, 1999.
12. Классификация – Machine Learning [Электронный ресурс] // [Режим доступа]: <http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D0%BB%D0%B0%D1%81%D1%81%D0%B8%D1%84%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D1%8F>
13. Лінгвістична змінна – Вікіпедія [Электронный ресурс] // Режим доступа: [https://uk.wikipedia.org/wiki/%D0%9B%D1%96%D0%BD%D0%B3%D0%B2%D1%96%D1%81%D1%82%D0%B8%D1%87%D0%BD%D0%B0\\_%D0%B7%D0%BC%D1%96%D0%BD%D0%BD%D0%B0](https://uk.wikipedia.org/wiki/%D0%9B%D1%96%D0%BD%D0%B3%D0%B2%D1%96%D1%81%D1%82%D0%B8%D1%87%D0%BD%D0%B0_%D0%B7%D0%BC%D1%96%D0%BD%D0%BD%D0%B0)
14. Марков А. А. Распространение закона больших чисел на величины, зависящие друг от друга. [Текст] / Марков А. А. // Известия физико-математического общества при Казанском университете. — 2-я серия. — Том 15. (1906) — С. 135–156.
15. Матвеев Л.П. Общая теория спорта и ее прикладные аспекты. [Текст] / Матвеев Л.П. // М.: ФГУП “Известия”, 2001.
16. Машинне навчання для прогнозування тенісу [Электронный ресурс] // [Режим доступа]: <http://it-ua.info/news/2016/08/09/mashinne-navchannya-dlya-prognozuvannya-tensu-chastina-2.html>
17. Машинное обучение для прогнозирования тенниса – Habrhabr [Электронный ресурс] // [Режим доступа]: <https://m.habr.com/post/307422/>
18. Метцлер П. Теннис [Текст] / Метцлер П. // М.: Агентство «ФАИР» , 1997. — 336 с.
19. Научные статьи в области искусственного интеллекта: Прогнозирование исходов спортивных игр методами нейросетевой кластеризации [Электронный ресурс] // Режим доступа: <http://neuronus.com/stat/207->

20. Нечітка логіка – Вікіпедія [Електронний ресурс] // Режим доступу: [https://uk.wikipedia.org/wiki/%D0%9D%D0%B5%D1%87%D1%96%D1%82%D0%BA%D0%B0\\_%D0%BB%D0%BE%D0%B3%D1%96%D0%BA%D0%B0](https://uk.wikipedia.org/wiki/%D0%9D%D0%B5%D1%87%D1%96%D1%82%D0%BA%D0%B0_%D0%BB%D0%BE%D0%B3%D1%96%D0%BA%D0%B0)
21. Офіційний сайт турнірів АРТworldtour [Електронний ресурс] // Режим доступу: [www.atpworldtour.com](http://www.atpworldtour.com)
22. Офіційний сайт турнірів Tennis-Data [Електронний ресурс] // Режим доступу: [www.tennis-data.co.uk](http://www.tennis-data.co.uk)
23. Платонов В.Н. Общая теория подготовки спортсменов в олимпийском спорте [Текст] / Платонов В.Н. // Киев, 1997.
24. Погодина С.В., Бекетов В.А. Основы общей теории спорта” [Текст] // Погодина С.В., Бекетов В.А. / Таврический национальный университет им. В.И. Вернадского, Симферополь 2011.
25. Програмне забезпечення OnCourt [Електронний ресурс] // Режим доступу: [http://www.oncourt.info/index\\_rus.html](http://www.oncourt.info/index_rus.html)
26. С. И. Мирошниченко. Цифровые приемники рентгеновских изображений [Текст] / С. И. Мирошниченко // Киев: издательство “Медицина Украины”, 2014. – 98 с.
27. Телеоптик [Електронний ресурс] // Режим доступу: <http://teleoptic-ltd.com>
28. Теніс – Вікіпедія [Електронний ресурс] // Режим доступу: <https://en.wikipedia.org/wiki/Tennis>
29. Технологія відслідковування тенісного м’яча HawkEye [Електронний ресурс] // Режим доступу: <http://www.hawkeyeinnovations.co.uk/>
30. Укроборонпром [Електронний ресурс] // Режим доступу: <http://ukroboronprom.com.ua/uk/>
31. Фестиваль інноваційних проєктів “Sikorsky Challenge” [Електронний ресурс] // Режим доступу: <https://www.sikorskychallenge.com/festival/>

32. A. M. Madurska. A Set-By-Set Analysis Method for Predicting the Outcome of Professional Singles Tennis Matches. [Текст] / A. M. Madurska // Technical report, Imperial College London, London, 2012.
33. A. Somboonphokkaphan. Tennis Winner Prediction based on Time-Series History with Neural Modeling. [Текст] / A. Somboonphokkaphan, S. Phimoltares, and C. Lursinsap // IMECS 2009: International Multi-Conference of Engineers and Computer Scientists, Vols I and II, I:127–132, 2009.
34. ATP World Tour: Tennis Statistics. [Электронный ресурс] // Режим доступа: <http://www.atpworldtour.com/en/players/alexandrdolgopolov/d801/player-stats>
35. Boeing [Электронный ресурс] // Режим доступа: <http://www.boeing.com>
36. Braga A. P.. Redes Neurais Artificiais [Текст] / Braga A. P., Carvalho A., Ludermir T. // Teoria e Aplicações. Rio de Janeiro, RJ, Brazil: LTC, 2005.
37. Bud Collins. Tennis encyklopedia [Текст] / Bud Collins and Zander Hollander // Visible Ink Press, 1998.
38. Christian Quidet. La fabuleuse histoire du tennis, [Текст] / Christian Quidet // Paris, Nathan, 1989.
39. Corral, J. Are differences in ranks good predictors for Grand Slam tennis matches? [Текст] / Corral, J., Prieto-Rorríguez, J. // International Journal of Forecasting, 26(3), 2010, 551–563.
40. Cybenko G. Approximation by superpositions of a sigmoidal function. [Текст] / Cybenko G. // Mathematics of Controls, Signals, and Systems, Springer Verlag, n. 2, p. 303-314, 1989.
41. F. J. G. M. Klaassen. Are Points in Tennis Independent and Identically Distributed? Evidence From a Dynamic Binary Panel Data Model. / F. J. G. M. Klaassen, J. R. Magnus // Journal of the American Statistical Association, 96:500-509, 2001.

42. Haykin S. Neural Networks [Текст] / Haykin S.// A Comprehensive Foundation. Upper Saddle River, NJ, USA: Prentice Hall, 1998.
43. J. Kelly. A new interpretation of information rate. [Текст] / J. Kelly // IRE Transactions on Information Theory, 2(3):917-926, 1956.
44. J.A. O'Malley. Probability Formulas and Statistical Analysis in Tennis. [Текст] / J.A. O'Malley // Journal of Quantitative Analysis in Sports, 4(2), 2008.
45. L. Prechelt. Early stopping - but when? In Neural Networks: Tricks of the Trade [Текст] / L. Prechelt // volume 1524 of LNCS, chapter 2, pages 55–69. Springer-Verlag, 1997.
46. Locke, E. A. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. [Текст] / Locke, E. A., Latham, G. P. // American Psychologist, 57(9), 2002, 705–717.
47. Loffing, F. The serve in professional men's tennis: Effects of players' handedness. [Текст] / Loffing, F., Hagemann, N., & Strauss, B. // International Journal of Performance Analysis in Sport, 9, 2009, 255–274.
48. M. A. Fernandes. Using Soft Computing Techniques for Prediction of Winners in Tennis Matches [Текст] / M. A. Fernandes // Federal Institute of Education, Science and Technology in Sergipe, Aracaju/SE, Brazil, 2017.
49. M. Bevc. Predicting the Outcome of Tennis Matches From Point-by-Point Data [Текст] / M. Bevc // School of Computing Science, Glasgow, 2015.
50. M. Sipko Machine Learning for the Prediction of Professional Tennis Matches. [Текст] / M. Sipko // Technical report, Imperial College London, London, 2015.
51. Magnus. Are Points in Tennis Independent and Identically Distributed? Evidence From a Dynamic Binary Panel Data Model [Текст] / Magnus, Klaassen // Journal of The American Statistical Association, Vol.96, pages 500-509, 2001.

52. Martin Hedges. The Concise Dictionary of Tennis. [Текст] / Martin Hedges // Mayflower Books Inc, 1978.
53. Neuroph Framework [Электронный ресурс] // [Режим доступа]: <http://neuroph.sourceforge.net/documentation.html>
54. Official Rules of Tennis [Электронный ресурс] // Режим доступа: <http://www.itftennis.com/officiating/rulebooks/rules-of-tennis.aspx>
55. Paul K. Newton. Monte Carlo Tennis: A Stochastic Markov Chain Model [Текст] / Paul K. Newton, Kamran Aslam // Journal of Quantitative Analysis in Sports, Berkeley Electronic Press, Vol. 5(3), 2009.
56. Peter Scholl. Richtig Tennis. [Текст] / Peter Scholl // München, BLV Buchverlag GmbH & Co., 5., neu bearb. Aufl., 2006.
57. Pinnacle [Электронный ресурс] // Режим доступа: <https://www.pinnacle.com/ru/>
58. Progresstech Ukraine [Электронный ресурс] // Режим доступа: <http://ptu.aero>
59. Rules of Tennis [Электронный ресурс] // Режим доступа: <http://library.thinkquest.org/C0125827/tennis.htm>.
60. S. Ma. Winning matches in Grand Slam men's singles: an analysis of player performance-related variables from 1991 to 2008. [Текст] / S. Ma, C. Liu, and Y. Tan // Journal of sports sciences, 31(11):1147–55, 2013.
61. S. R. Clarke. Using official ratings to simulate major tennis tournaments. [Текст] / S. R. Clarke, D. Dyte // International Transactions in Operational Research, 7(6):585–594, 2000.
62. Sugeno M. Industrial Applications of Fuzzy Control. [Текст] / Sugeno M. // New York, NY, USA: Elsevier Science Pub. Co., 1985.

63. T. Barnett. Combining player statistics to predict outcomes of tennis matches. [Текст] / T. Barnett, S. R. Clarke // IMA Journal of Management Mathematics, 16:113–120, 2005.
64. T. Barnett. The Mathematics of Tennis [Текст] / T. Barnett, A. Brown // Strategic Games, 2012, 170 p.
65. T. Kohonen. Learning Vector Quantization [Текст] / T. Kohonen // Neural Networks, 1 (suppl 1), 303, 1988.
66. Tennis [Електронний ресурс] // Режим доступу: <http://www.tradgames.org.uk/games/Tennis.htm>.
67. Tennis Court Dimensions [Електронний ресурс] // Режим доступу: <http://www.sportsknowhow.com/tennis/dimensions/tennis-court-dimensions.html>
68. W. J. Knottenbelt. A common-opponent stochastic model for predicting the outcome of professional tennis matches. [Текст] / W. J. Knottenbelt, D. Spanias, and A. M. Madurska // Computers and Mathematics with Applications, 64:3820–3827, 2012.
69. Zadeh L. Fuzzy Sets. [Текст] / Zadeh L. // Information and Control, n. 8: p. 338-353, 1965.
70. Дідківська В.А. Модифікована методологія прогнозування результатів спортивних змагань [Текст]: матеріали 4-ї міжнародної науково-практичної конференції “Актуальні питання сучасної науки”, Київ, 16-17 травня 2018 р./ Дідківська В.А., Гавриленко О.В.
71. Дідківська В.А. Підхід до прогнозування результатів спортивних змагань [Текст]: матеріали науково практичної конференції “Інформатика та обчислювальна техніка ІОТ-2018”, Київ: НТУУ “КПІ ім. Сікорського”, 23-24 квітня 2018 р. / Дідківська В.А., Гавриленко О.В.

72. Гавриленко О.В. Марківські процеси у прогнозуванні результатів тенісних матчів [Текст] / Гавриленко О.В., Дідківська В.А. “Управління проектами, системний аналіз та логістики”. – Київ, 2018 . – Серія “Технічні науки”
73. Гавриленко О.В. Марківські процеси у прогнозуванні результатів тенісних матчів [Текст]: матеріали 8-ї міжнародної науково-технічної конференції “Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління”, Харків, 26-27 квітня 2018 р. /Гавриленко О.В., Дідківська В.А. – 108 с. – с.67-68



## **ДОДАТОК А**

### **Сертифікат навчання у Стартап Школі “Sikorsky Challenge”**



Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Навчально-методичний комплекс «Інститут післядипломної освіти»  
Training and Methodical Complex "Institute for Continuing Education"

«Be Next IT» — Відкрита інноваційна стартап школа-інкубатор (Ізраїль)  
"Be Next IT" — Open Innovation Startup School-Incubator (Israel)

# СЕРТИФІКАТ CERTIFICATE

засвідчує, що  
this is to certify that

TPC № 5002

**Дідківська Валерія**  
**Valeriia Didkivska**

з 13 лютого по 20 квітня 2017 року  
from February 13 to April 20, 2017

Пройшов(ла) навчання у Стартап Школі «Sikorsky Challenge»  
за програмою

Took a course of training in the Startup School  
"Sikorsky Challenge" under the program

**«Вступ до інноваційного підприємництва  
та практика запуску стартапа»**

**"Introduction to Innovative Entrepreneurship  
and Practice of Launching a Startup"**

Директор НМК «Інститут  
післядипломної освіти»  
Director of TMC "Institute  
for Continuing Education"

Директор «Be Next IT»  
Director of "Be Next IT"

I. Малюкова  
I. Maliukova

I. Пеер  
I. Peer



## **ДОДАТОК Б Графічний матеріал**

## **ПЛАКАТ 1 Фрагмент діаграми класів**

## **ПЛАКАТ 2 Марківський ланцюг для ситуації сету і тай-брейку**

### **ПЛАКАТ 3 Діаграма бази даних**

## ПЛАКАТ 4 Діаграма Use-case

**ПЛАКАТ 5 Схеми роботи алгоритму в загальному вигляді**



## **ПЛАКАТ 6 Система виводу задачі**

## **ПЛАКАТ 7 Екранні форми початку роботи програми**

## **ПЛАКАТ 8 Екранні форми процесу роботи програми**